



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47, 16/18, G01N 33/53</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/53061</b> <b>(43) International Publication Date:</b> 26 November 1998 (26.11.98)																	
<b>(21) International Application Number:</b> PCT/AU98/00380 <b>(22) International Filing Date:</b> 22 May 1998 (22.05.98) <b>(30) Priority Data:</b> <table border="0"> <tr> <td>PO 6972</td> <td>23 May 1997 (23.05.97)</td> <td>AU</td> </tr> <tr> <td>PO 6973</td> <td>23 May 1997 (23.05.97)</td> <td>AU</td> </tr> <tr> <td>PO 6974</td> <td>23 May 1997 (23.05.97)</td> <td>AU</td> </tr> <tr> <td>PP 1458</td> <td>22 January 1998 (22.01.98)</td> <td>AU</td> </tr> <tr> <td>PP 1459</td> <td>22 January 1998 (22.01.98)</td> <td>AU</td> </tr> <tr> <td>PP 1460</td> <td>22 January 1998 (22.01.98)</td> <td>AU</td> </tr> </table> <b>(71) Applicant (for all designated States except US):</b> THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH [AU/AU]; 300 Herston Road, Brisbane, QLD 4029 (AU). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> HAYWARD, Nicholas [AU/AU]; 13 Prince Street, Paddington, QLD 4064 (AU). SILINS, Ginters [AU/AU]; 35 Eppalong Street, The Gap, QLD 4061 (AU). GRIMMOND, Sean [AU/GB]; Medical Research Council, Harwell, Didcot, Oxfordshire OX11 0RD (GB). GARTSIDE, Michael [AU/AU]; 19 Thomas Street, Camp Hill, QLD 4152 (AU). HANCOCK, John [AU/AU]; 141 Airley Road, Pullenvale, QLD 4069 (AU).	PO 6972	23 May 1997 (23.05.97)	AU	PO 6973	23 May 1997 (23.05.97)	AU	PO 6974	23 May 1997 (23.05.97)	AU	PP 1458	22 January 1998 (22.01.98)	AU	PP 1459	22 January 1998 (22.01.98)	AU	PP 1460	22 January 1998 (22.01.98)	AU	<b>(74) Agents:</b> HUGHES, E., John, L. et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU). <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>
PO 6972	23 May 1997 (23.05.97)	AU																	
PO 6973	23 May 1997 (23.05.97)	AU																	
PO 6974	23 May 1997 (23.05.97)	AU																	
PP 1458	22 January 1998 (22.01.98)	AU																	
PP 1459	22 January 1998 (22.01.98)	AU																	
PP 1460	22 January 1998 (22.01.98)	AU																	
<b>(54) Title:</b> THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN  <b>(57) Abstract</b> <p>The present invention relates generally to three novel human genes with gene regulatory function. These genes encode a zinc finger protein, a guanine nucleotide exchange protein and a heat shock protein or heat shock binding protein. The invention includes derivatives and mammalian animal, insect, nematodes, avian and microbial homologues of these genes. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.</p>																			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

## FIELD OF THE INVENTION

5 The present invention relates generally to a novel human gene and its derivatives and to mammalian, animal, insect, nematodes, avian and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

10

## BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

15

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis and in conventional pharmaceutical preparations as well as in gene and protein replacement therapies.

20

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. Molecules of particular interest targeted by the inventors were gene regulators including regulatory proteins, signal transducers and heat shock proteins.

25

Gene expression generally requires interaction between a regulatory protein and an appropriate recognition sequence of a target gene. Regulatory proteins comprise in many cases a domain or motif which facilitates binding to DNA. One particular motif comprises small sequence units repeated in tandem with each unit folded about a zinc atom to form separate structural domains.

30 This motif is now referred to as a zinc finger domain. Such a domain is generally defined by the number of cysteine (C) and histidine (H) residues.

In addition, knowledge of cellular interaction in the control of cell proliferation is essential in the rational design of specific therapeutic strategies aimed at controlling proliferative disorders. Such proliferative disorders including a range of cancers, inflammatory conditions and atherosclerosis. An important aspect of cellular interaction is in signal transduction *via* receptors  
5 to intracellular transducers. One key signal transducer is Ras which couples the receptors for diverse extracellular signals to different effectors. Ras directly activates the downstream kinase Raf which in turn induces the mitogen activated protein kinase (MAPK) cascade.

Another regulatory mechanism involves heat shock proteins. The *Escherichia coli* heat shock  
10 protein, DnaJ, is the founding member of a family of proteins which are associated with protein folding, protein complex assembly and transit through subcellular components.

Prokaryotic and eukaryotic DnaJ homologues have a modular organisation consisting of a J domain, a glycine-rich spacer, CXXCXGXG [SEQ ID NO:1] repeats and a C-terminal region  
15 with no obvious sequence features, as well as additional sequences for protein targeting. The J domain is anticipated to mediate interaction with heat shock 70 proteins (Hsp70) and consists of some 70 amino acids, frequently located at the N-terminus of the protein.

In accordance with the present invention, a genes have been identified from the human genome  
20 which encodes proteins having a regulatory role. One gene, in accordance with the present invention encodes a protein with an N-terminal region resembling a zinc-finger domain of a novel type. Another gene encodes a protein involved in guanine nucleotide exchange factor (GEF) signalling pathways. Yet another gene encodes a protein which is a heat shock protein or heat shock-like protein which may have a role in tumour suppression.

25

#### SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a  
30 stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.



Sequence identity numbers (SEQ ID NOs.) for nucleotide and amino acid sequences referred to in the subject specification are defined after the bibliography. A summary of SEQ ID NOs. is also given in Table 1.

- 5 One aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
- 10 Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an  $(\text{HC}_3)_2$  type.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule  
15 comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence  
20 of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:2 defines the gene, *mcg4*. This gene encodes  
25 a product, MCG4, having an amino acid sequence set forth in SEQ ID NO:3.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg4* gene portion, which *mcg4* gene portion is capable of encoding an MCG4 polypeptide or a  
30 functional or immunologically interactive derivative thereof.

- 4 -

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion  
5 and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or  
10 multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological  
15 sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

A further aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an  
20 amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

25

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence  
30 of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:4 or 6 defines the gene, *mcg7*. This gene encodes a product, MCG7, having an amino acid sequence set forth in SEQ ID NO:5 or 7.

5

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an MCG7 polypeptide or a functional or immunologically interactive derivative thereof.

10

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock binding protein or a derivative thereof.

Another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 5 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 41°C to the nucleotide sequence set forth in (i), (ii) or (iii).

10

The nucleotide sequence set forth in SEQ ID NO:8 defines the gene, *mcg18*. This gene encodes a product, MCG18, having an amino acid sequence set forth in SEQ ID NO:7.

Even yet another aspect of the present invention provides a genetic construct comprising a vector  
15 portion and an animal, more particularly a mammalian and even more particularly a human *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition  
20 caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

25

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

30

Another aspect of the present invention contemplates a method for detecting MCG18 or a

- 7 -

derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

5

A summary of SEQ ID Nos. referred to in the subject specification is shown in Table 1.

**TABLE 1**  
**SUMMARY OF SEQ ID Nos.**

5	SEQ ID NO.	DESCRIPTION
	1	amino acid repeat sequence in DnaJ homologues
	2	Nucleotide sequence of <i>mcg4</i>
	3	amino acid sequence of MCG4
	4	nucleotide sequence of <i>mcg7</i>
10	5	amino acid sequence of MCG7
	6	nucleotide sequence of <i>mcg7</i> within exon of nucleotides 183-288
	7	amino acid sequence of MCG7 within exon of nucleotide 183-288
	8	nucleotide sequence of <i>mcg18</i>
	9	amino acid sequence of MCG18
15	10-18	amino acid sequence identified using BESTFIT
	19	sequence of pGEX and <i>mcg7</i> junction
	20	sequence of pGEX and <i>mcg7</i> junction
	21	nucleotide sequence of <i>myc</i> -tag/ <i>mcg7</i> junction
	22	amino acid sequence corresponding to SEQ ID NO:21
20	23	nucleotide sequence of pGEX and <i>mcg7</i> junction
	24	amino acid sequence corresponding to SEQ ID NO:23
	25-36	<i>mcg7</i> -specific oligonucleotide
	37-45	<i>mcg18</i> -specific oligonucleotide

25 Single and three letter abbreviations for amino acid residues are shown in Table 2.

- 9 -

TABLE 2

Amino Acid	Three-letter Abbreviation	One-letter Symbol
5		
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
10 Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
15 Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
20 Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
25 Valine	Val	V
Any residue	Xaa	X

**BRIEF DESCRIPTION OF THE FIGURES**

**Figure 1** is a representation of the nucleotide sequence [SEQ ID NO:2] and corresponding amino acid sequence [SEQ ID NO:3] of *mcg4*.

5

**Figure 2** is a representation of the alignment of the human MCG4 amino acid sequence with a translation of a partial murine expressed sequence tag (EST).

**Figure 3** is a representation of the alignment of the human MCG4 amino acid sequence with a  
10 translation of a partial nematode EST.

**Figure 4** is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn represent zinc atoms.

15

**Figure 5** is a representation of sensitive sequence homology search of related cysteine-containing motifs in another *Caenorhabditis elegans* protein.

**Figure 6** is a representation showing that a related cysteine containing motif is present in the  
20 GATA-binding transcription factor from *Saccharomyces pombe*.

**Figure 7** is a Northern blot showing expression of *mcg4* in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15µg total RNA derived from various human cancer cell lines. Lanes 1-5, respectively, contain RNA from H69 lung  
25 carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

**Figure 8** is a representation of a partial alignment of *mcg4* with human ESTs AA074703 and AA134788.

30

**Figure 9** is a representation of the partial nucleotide sequence alignment between a human



- 11 -

(W32939) and mouse (AA242159) *mcg4*-like EST in the putative 5' UTR of the *mcg4* cDNA. The putative initiation codon is underlined and the region upstream represents 5' UTR.

**Figure 10** is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

**Figure 11** is a diagrammatic representation of the domains of MCG4

zinc finger consensus: CX<sub>2</sub>HX<sub>4</sub>CX<sub>2</sub>CX<sub>4</sub>HX<sub>2</sub>CX<sub>17</sub>CX<sub>2</sub>CX<sub>18</sub>HX<sub>2</sub>CX<sub>18</sub>CX<sub>2</sub>C

acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged

10 basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged

leucine zipper domain consensus: LX<sub>6</sub>LX<sub>6</sub>RX<sub>6</sub>LX<sub>6</sub>L

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa261) LX<sub>6</sub>LXLX<sub>6</sub>LXLX<sub>6</sub>L (aa 286).

15 **Figure 12** is a representation showing similarity of MCG7 with GEFs of various organisms.

**Figure 13(a)** is a representation of the nucleotide sequence [SEQ ID NO:4] and corresponding amino acid sequence [SEQ ID NO:5] of *mcg7*. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

20

**Figure 13(b)** is a representation of the partial nucleotide sequence [SEQ ID NO:6] and corresponding amino acid sequence [SEQ ID NO:7] of *mcg7* but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon  
25 of nucleotides 183-288.

**Figure 14** is a representation showing a comparison between MCG7 and a homologue from *Caenorhabditis elegans* using the BESTFIT algorithm. In the figure, the following sequences are underlined:

30

EF-Hand= PROSITE DATABASE NO. PD0C00018

- 12 -

1a nematode DVDEEDEVEDIEF [SEQ ID NO:10]  
 1b human DVDGDGHISQEEF [SEQ ID NO:11]  
     nematode DHDRDGFISQEEF [SEQ ID NO:12]  
 1c human DQNQDGCISREEM [SEQ ID NO:13]  
 5     nematode DVDMDGQISKDEL [SEQ ID NO:14]

**GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B**

2     human     HFVHVAEKLLQLQNFTLMAVVGGLSHSSISRLKETH[SEQ ID NO:15]  
           nematode   KFBVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTY  
 10                   [SEQ ID NO:16]

**DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379**

3     human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC  
                   [SEQ ID NO:17]  
 15    nematode HNFHETTFLTPPTCNHCNKLLWGILRQGFCKDCGLAVHSCCKSNAVAEC  
                   [SEQ ID NO:18]

**Figure 15** is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine *mcg7* cDNA (GenBank Acc. No. W71787 and AA237373). The putative  
 20 initiation codon is underlined. The murine sequence represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

25 **Figure 16** is a representation of further 5' nucleotide and corresponding amino acid sequence for human *mcg7*. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Fig. 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.

30 **Figure 17** is a graphical representation of a GDP release assay. □ Experiment #1 (mean of duplicates). ◇ Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols

of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 mins.

Estimated reaction constants:

$K_m = 2.1\mu\text{M}$ ,  $V_{max} = 37\text{pMol}/6\text{min}/36\text{pMol}$  [Expt#1]

5  $K_m = 1.5\mu\text{M}$ ,  $V_{max} = 30.3\text{pMol}/6\text{ min}/36\text{pMol}$  [Expt#2]

**Figure 18** depicts various recombinant plasmids containing partial or full-length *mcg7*.

**Figure 19** is a representation of the nucleotide sequence [SEQ ID NO:8] and corresponding  
10 amino acid sequence [SEQ ID NO:9] of *mcg18*.

**Figure 20** is a representation showing that MCG18 has partial homology to *E. coli* DnaJ.

**Figure 21** is a representation showing that MCG18 has homology to two *Caenorhabditis elegans*  
15 proteins.

**Figure 22** is a representation showing that MCG18 has homology to a *Saccharomyces pombe* protein.

20 **Figure 23** is a representation showing homology of MCG18 to a *Drosophila virilis* protein.

**Figure 24** is a representation showing homology of MCG18 to human DnaJ proteins HDJ-2/HSDJ, HDJ-1/HSP40 and HSJ1.

25 **Figure 25** is a representation of the nucleotide and corresponding amino acid sequence of murine *mcg18*.

**Figure 26** is a representation of homology between human and murine MCG18.

30 **Figure 27** depicts nucleotide sequences corresponding to the 5' untranslated region of human *mcg18*.

**Figure 28** depicts a Northern blot showing expression of *mcg18* transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain 15 $\mu$ g RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

**DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

The present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having  
5 homology to a regulator of gene expression or a derivative of said gene regulator.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an  $(\text{HC}_3)_2$   
10 type.

Still more particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- 15 (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C  
20 to the nucleotide sequence set forth in (i), (ii) or (iii).

The present invention also provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.  
25

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- 30 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;

- 16 -

- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

5

Another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock-binding protein or a derivative thereof.

10

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 15 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

20

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1%  
25 v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M  
30 to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least

- 17 -

about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

The present invention extends to nucleic acid molecules with percentage similarities of approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

15 The nucleic acid molecule of the present invention defined by SEQ ID NO:2 is hereinafter referred to as constituting the "*mcg4*" gene. The protein encoded by *mcg4* is referred to herein as "MCG4" and has an amino acid sequence set forth in SEQ ID NO:3. The *mcg4* gene is proposed to encode, in accordance with the present invention, a regulator of gene expression and comprises a novel zinc finger domain,  $(HC_3)_2$ . A regulator of gene expression includes a transcription factor. Regulation may be at the level of nucleic acid:protein or protein:protein interaction.

The nucleic acid molecule of the present invention defined by SEQ ID NO:4 or 6 is hereinafter referred to as constituting the "*mcg7*" gene. The protein encoded by *mcg7* is referred to herein as "MCG7" and has an amino acid sequence set forth in SEQ ID NO:5 or 7 and is involved in signal transduction. The difference in the nucleotide and amino acid sequence is due to the presence or absence of an exon at nucleotides 183-288.

The nucleic acid molecule of the present invention defined by SEQ ID NO:8 is hereinafter referred to as constituting the "*mcg18*" gene. The protein encoded by *mcg18* is referred to herein as "MCG18" and comprises the amino acid set forth in SEQ ID NO:9.

The present invention extends to the naturally occurring genomic *mcg4*, *mcg7* and *mcg18* nucleotide sequences or corresponding cDNA sequences or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG4, MCG7 or MCG8 or the corresponding genetic sequences. Derivatives  
5 also include single or multiple amino acid substitutions, deletions and/or additions to MCG4, MCG7 or MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to *mcg4*, *mcg7* or *mcg18*. "Additions" to the amino acid or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG4" or "*mcg4*", "MCG7" or "*mcg7*" or "MCG8" or *mcg18*" includes reference to  
10 all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG4, MCG7 or MCG18.

The *mcg4*, *mcg7* and *mcg18* of the present invention are particularly exemplified herein from humans and in particular from human chromosome 11q13.

15

The present invention extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), reptiles, birds (eg. chickens, ducks, geese, parrots), insects, nematodes, eukaryotic microorganisms and captive wild animals (eg.  
20 deer, foxes, kangaroos). Reference herein to *mcg4* and *mcg18* or their respective proteins MCG4, MCG7 and MCG18 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic  
25 acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic  
30 molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or



- 19 -

both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

- 5 Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg4* gene portion, which *mcg4* gene portion is capable of encoding an MCG4 polypeptide or a functional or immunologically interactive derivative thereof.
- 10 Preferably, the *mcg4* gene portion of the genetic construct is operably linked to a promoter in the vector such that said promoter is capable of directing expression of said *mcg4* gene portion in an appropriate cell.

In addition, the *mcg4* gene portion of the genetic construct may comprise all or part of the gene  
15 fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

20

It is proposed in accordance with the present invention that MCG4 is a transcription factor involved in gene regulation. Mutations in *mcg4* may result in aberrations in gene regulation leading to the development of or a propensity to develop various types of cancer. In this regard, although not wishing to limit the present invention to any one hypothesis or mode of action, it  
25 is proposed that *mcg4* or its expression product may be involved in the tissue-specific or temporal regulation of particular genes.

A deletion or aberration in the *mcg4* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a  
30 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may

- 20 -

be determined by assaying for aberrations in the parents and/or proband of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting  
5 a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

10

Another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an *mcg7* polypeptide or a functional or immunologically interactive derivative thereof.

15

Preferably, the *mcg7* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg7* gene portion in an appropriate cell.

20 In addition, the *mcg7* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells  
25 comprising same.

It is proposed in accordance with the present invention that MCG7 is a GEF involved in signal transduction. Mutations in *mcg7* or MCG7 may result in defective control of cell proliferation leading to the development of or a propensity to develop various types of cancer.

30

A deletion or aberration in the *mcg7* gene may also be important in the detection of cancer or

a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents of a subject under investigation.

5

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide  
10 substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Yet another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human  
15 *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg18* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg18* gene portion  
20 in an appropriate cell.

In addition, the *mcg18* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

25

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG18 is a transcription factor  
30 involved in protein folding, protein complex assembly and transit through subcellular compartments. MCG18 may also have a role in tumour suppression. Thus mutations in *mcg18*

may result in the development of or a propensity to develop various types of cancer.

A deletion or aberration in the *mcg18* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a  
5 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents and/or proband of the subject under investigation.

10 According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other  
aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or  
15 a propensity to develop said condition.

The nucleotide substitutions, additions or deletions may be detected by any convenient means including nucleotide sequencing, restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), oligonucleotide hybridization and single stranded conformation  
20 polymorphism analysis (SSCP) amongst many others. An aberration includes modification to existing nucleotides such as to modify glycosylation signal amongst other effects.

In an alternative method, aberrations in the *mcg4*, *mcg7* and *mcg18* genes are detected by screening for mutations in MCG4, MCG7 and MCG18, respectively.

25

A mutation in MCG4, MCG7 or MCG18 may be a single or multiple amino acid substitution, addition and/or deletion. The mutation in *mcg4*, *mcg7* or *mcg18* may also result in either no translation product being produced or a product in truncated form. A mutant may also be an altered glycosylation pattern or the introduction of side chain modifications to amino acid  
30 residues.

- 23 -

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in *mcg4*, *mcg7* or *mcg18* said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4, MCG7 or MCG18 wherein the presence of such a mutation is indicative of or a propensity to  
5 develop said condition.

A particularly convenient means of detecting a mutation in MCG4, MCG7 or MCG18 is by use of antibodies.

- 10 Accordingly another aspect of the present invention is directed to antibodies to MCG4, MCG7 or MCG18 and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG4, MCG7 or MCG18 or may be specifically raised to MCG4, MCG7 or MCG18 or derivatives thereof. In the case of the latter, MCG4, MCG7 or MCG18 or their derivatives may first need to be associated with a carrier molecule.
- 15 The antibodies to MCG4, MCG7 or MCG18 of the present invention are particularly useful as diagnostic agents.

- For example, antibodies to MCG4, MCG7 or MCG18 and their derivatives can be used to screen for wild-type MCG4, MCG7 or MCG18 or for mutated MCG4, MCG7 or MCG18 molecules.
- 20 The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG4, MCG7 or MCG18 levels or the presence of wild-type MCG4, MCG7 or MCG18 may be important for diagnosis of certain cancers or a predisposition for development of cancers or for monitoring  
25 certain therapeutic protocols.

- As stated above antibodies to MCG4, MCG7 or MCG18 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to  
30 antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

- 24 -

For example, specific antibodies can be used to screen for wild-type MCG4, MCG7 or MCG18 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for example, as a means for screening for levels of MCG4, MCG7 or MCG18 in a cell extract or other biological fluid or purifying MCG4, MCG7 or MCG18 made by recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG4, MCG7 or MCG18 or to a specific mutant phenotype or to a deleted or otherwise altered region.

15

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG4, MCG7 or MCG18 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal or bird with an effective amount of MCG4, MCG7 or MCG18 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoabsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

25

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

30

Another aspect of the present invention contemplates a method for detecting MCG4, MCG7 or MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4, MCG7 or MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4, MCG7 or  
5 MCG18 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

The presence of MCG4, MCG7 or MCG18 may be accomplished in a number of ways such as  
10 by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4,424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

15

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into  
20 contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is  
25 determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the  
30 art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG4, MCG7 or MCG18 including cell extract

or tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the MCG4, MCG7  
5 or MCG18 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding  
10 processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 37°C) to allow binding of any subunit present in the  
15 antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

20 An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-  
25 first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-  
30 bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide



- 27 -

containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled  
5 artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a  
10 fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated,  
15 usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically  
20 coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the  
25 unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

30

As stated above, the present invention extends to genetic constructs capable of encoding MCG4,

MCG7 or MCG18 or functional derivatives thereof. Such genetic constructs are also contemplated to be useful in modulating expression of specific genes in which *mcg4*, *mcg7* or *mcg18* is involved in tissue-specific or temporal regulation.

- 5 Accordingly, another aspect of the present invention is directed to a genetic construct comprising a nucleotide sequence encoding a peptide, polypeptide or protein and *mcg4*, *mcg7* or *mcg18* or a functional derivative or homologue thereof capable of modulating the expression of said nucleotide sequence.
- 10 As stated above, MCG18 is proposed to have a role in tumour suppression. Accordingly, it is further proposed in accordance with the present invention to use recombinant MCG18 in pharmaceutical preparations for treating arresting or otherwise ameliorating the effects of certain cancers.
- 15 Accordingly, another aspect of the present invention contemplates a method for treating, arresting or otherwise ameliorating the effects of a cancer in an animal or bird, said method comprising administering to said animal or bird an effective amount of MCG18 or a functional derivative thereof for a time and under conditions sufficient to treat, arrest or otherwise ameliorate the effects of said cancer.
- 20
- The present invention, therefore, contemplates a pharmaceutical composition comprising MCG18 or a derivative thereof or a modulator of *mcg18* expression or MCG18 activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to hereinafter as the "active ingredients". The active ingredients may also include anti-cancer
- 25 agents or agents which facilitate actions of MCG18.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be

30 preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier may be a solvent medium containing, for example, water, ethanol, polyol (for example,

- 29 -

glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin and by the use of surfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, 5 chlorobutanol, phenol, sorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

- 10 Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired 15 ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with 20 the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 25 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions is such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1  $\mu$ g and 2000 mg of active compound.

- 30 The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium

phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of  
5 the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form  
10 should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

The present invention also extends to forms suitable for topical application such as creams,  
15 lotions and gels.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known  
20 in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease  
25 of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the  
30 unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the

treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in  
5 effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as  
hereinbefore disclosed. A unit dosage form can, for example, contain the principal active  
compound in amounts ranging from 0.5  $\mu\text{g}$  to about 2000 mg. Expressed in proportions, the  
active compound is generally present in from about 0.5  $\mu\text{g}$  to about 2000 mg/ml of carrier. In  
the case of compositions containing supplementary active ingredients, the dosages are  
10 determined by reference to the usual dose and manner of administration of the said ingredients.

Effective amounts contemplated by the present invention include those amounts effective to  
ameliorate a condition. For example, it is envisaged that effective amounts would range from  
about 0.001  $\mu\text{g/kg}$  body weight to about 100 mg/kg body weight. Alternatively, effective  
15 amounts of about 0.01  $\mu\text{g/kg}$  body weight to about 10 mg/kg body weight or even 0.1  $\mu\text{g/kg}$   
body weight to about 1 mg/kg body weight. Administration may be per minute, hour, day, week,  
month or year or may only be a once off administration.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable  
20 of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating  
*mcg18* expression or MCG18 activity. The vector may, for example, be a viral vector.

As stated above, the present invention further contemplates a range of derivatives of MCG18.  
Derivatives include fragments, parts, portions, mutants, homologues and analogues of the  
25 MCG18 polypeptide and corresponding genetic sequence. Derivatives also include single or  
multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple  
nucleotide substitutions, deletions and/or additions to the genetic sequence encoding MCG18.  
"Additions" to amino acid sequences or nucleotide sequences include fusions with other  
peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to  
30 "MCG18" includes reference to all derivatives thereof including functional derivatives or MCG18  
immunologically interactive derivatives.

Analogues of MCG18 contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

5

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with  $\text{NaBH}_4$ ; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups  
10 with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with  $\text{NaBH}_4$ .

The guanidine group of arginine residues may be modified by the formation of heterocyclic  
15 condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

20 Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and  
25 other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides.

Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form  
30 a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis  
5 include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 3.

TABLE 3

Non-conventional amino acid	Code	Non-conventional amino acid	Code
5 $\alpha$ -aminobutyric acid	Abu	L-N-methylalanine	Nmala
$\alpha$ -amino- $\alpha$ -methylbutyrate	Mgab	L-N-methylarginine	Nmarg
aminocyclopropane- carboxylate	Cpro	L-N-methylasparagine	Nmasn
		L-N-methylaspartic acid	Nmasp
10 aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl- carboxylate	Norb	L-N-methylglutamine	Nmgln
		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisoleucine	Nmile
15 D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
20 D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
25 D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
30 D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva



	D-tyrosine	Dtyr	$\alpha$ -methyl-aminoisobutyrate	Maib
	D-valine	Dval	$\alpha$ -methyl- $\gamma$ -aminobutyrate	Mgab
	D- $\alpha$ -methylalanine	Dmala	$\alpha$ -methylcyclohexylalanine	Mchexa
	D- $\alpha$ -methylarginine	Dmarg	$\alpha$ -methylcyclopentylalanine	Mcpen
5	D- $\alpha$ -methylasparagine	Dmasn	$\alpha$ -methyl- $\alpha$ -naphthylalanine	Manap
	D- $\alpha$ -methylaspartate	Dmasp	$\alpha$ -methylpenicillamine	Mpen
	D- $\alpha$ -methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D- $\alpha$ -methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D- $\alpha$ -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D- $\alpha$ -methylisoleucine	Dmile	N-amino- $\alpha$ -methylbutyrate	Nmaabu
	D- $\alpha$ -methylleucine	Dmleu	$\alpha$ -naphthylalanine	Anap
	D- $\alpha$ -methyllysine	Dmlys	N-benzylglycine	Nphe
	D- $\alpha$ -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D- $\alpha$ -methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D- $\alpha$ -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D- $\alpha$ -methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D- $\alpha$ -methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D- $\alpha$ -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D- $\alpha$ -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D- $\alpha$ -methyltyrosine	Dmtty	N-cyclodecylglycine	Ncdec
	D- $\alpha$ -methylvaline	Dmval	N-cyclododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmglu	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl)glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

	D-N-methyllysine	Dnmlys	N-methyl- $\gamma$ -aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmt
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
5	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyl- $\alpha$ -naphthylalanine	Nmanap
10	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	$\gamma$ -aminobutyric acid	Gabu	N-( <i>p</i> -hydroxyphenyl)glycine	Nhtyr
	L- <i>t</i> -butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L- $\alpha$ -methylalanine	Mala
15	L- $\alpha$ -methylarginine	Marg	L- $\alpha$ -methylasparagine	Masn
	L- $\alpha$ -methylaspartate	Masp	L- $\alpha$ -methyl- <i>t</i> -butylglycine	Mtbug
	L- $\alpha$ -methylcysteine	Mcys	L-methylethylglycine	Metg
	L- $\alpha$ -methylglutamine	Mgln	L- $\alpha$ -methylglutamate	Mglu
	L- $\alpha$ -methylhistidine	Mhis	L- $\alpha$ -methylhomophenylalanine	Mhphe
20	L- $\alpha$ -methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L- $\alpha$ -methylleucine	Mleu	L- $\alpha$ -methyllysine	Mlys
	L- $\alpha$ -methylmethionine	Mmet	L- $\alpha$ -methylnorleucine	Mnle
	L- $\alpha$ -methylnorvaline	Mnva	L- $\alpha$ -methylornithine	Morn
	L- $\alpha$ -methylphenylalanine	Mphe	L- $\alpha$ -methylproline	Mpro
25	L- $\alpha$ -methylserine	Mser	L- $\alpha$ -methylthreonine	Mthr
	L- $\alpha$ -methyltryptophan	Mtrp	L- $\alpha$ -methyltyrosine	Mtyr

- 37 -

L- $\alpha$ -methylvaline	Mval	L-N-methylhomophenylalanine	Nmhph
N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe
1-carboxy-1-(2,2-diphenyl- 5 ethylamino)cyclopropane	Nmbc		

---

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having  $(CH_2)_n$  spacer groups with  $n=1$  to  $n=6$ ,  
 10 glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of  $C_\alpha$  and  $N_\epsilon$ -methylamino acids, introduction of double bonds between  $C_\alpha$  and  $C_\beta$  atoms of amino acids and  
 15 the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

Such analogues also apply in respect of MCG4 and MCG7.

20

The present invention further contemplates chemical analogues of MCG18 capable of acting as antagonists or agonists of MCG18 or which can act as functional analogues of MCG18. Chemical analogues may not necessarily be derived from MCG18 but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to  
 25 mimic certain physiochemical properties of MCG18. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

The identification of MCG18 permits the generation of a range of therapeutic molecules capable of modulating expression of MCG18 or modulating the activity of MCG18. Modulators  
 30 contemplated by the present invention includes agonists and antagonists of MCG18 expression. Antagonists of MCG18 expression include antisense molecules, ribozymes and co-suppression

molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of MCG18 include molecules which overcome any negative regulatory mechanism. Antagonists of MCG18 include antibodies and inhibitor peptide fragments.

5

These types of modifications may be important to stabilise MCG18 if administered to an individual or for use as a diagnostic reagent.

Other derivatives contemplated by the present invention include a range of glycosylation variants  
10 from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

Another embodiment of the present invention contemplates a method for modulating expression  
15 of MCG18 in a human, said method comprising contacting the *mcg18* gene encoding MCG18 with an effective amount of a modulator of *mcg18* expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression of *mcg18*. For example, a nucleic acid molecule encoding MCG18 or a derivative thereof may be introduced into a cell to facilitate protection of that cell from becoming cancerous.

20

Another aspect of the present invention contemplates a method of modulating activity of MCG18 in a human, said method comprising administering to said mammal a modulating effective amount of a molecule for a time and under conditions sufficient to increase or decrease MCG18 activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative  
25 of MCG18 or a chemical analogue or truncation mutant of MCG18.

The present invention is further described with reference to the following non-limiting Examples.

- 39 -

**EXAMPLE 1**

A human gene (designated *mcg4*) was identified on chromosome 11q13 that on the basis of sequence homology is predicted to encode a putative transcription factor of 310 amino acids  
5 (Fig. 1). *mcg4* is transcribed in several different cell lines (Fig. 7).

**EXAMPLE 2**

The expressed sequence tag (EST) database contains partial sequence data for the murine (Fig.  
10 2) and nematode (Fig. 3) homologues of *mcg4*.

**EXAMPLE 3**

MCG4 contains a sequence of cysteine residues within the N-terminal region of the protein that  
15 resembles zinc-finger binding domains of a novel type, ie.  $(HC_3)_2$  [Fig. 4].

**EXAMPLE 4**

Sensitive sequence homology searches reveal that related cysteine-containing motifs are present  
20 in another *C. elegans* protein (Fig. 5) as well as the GATA-binding transcription factor from *S. pombe* (Fig. 6).

**EXAMPLE 5**

25 *mcg4* will have commercial value due to its likelihood of encoding a novel transcription factor that is highly conserved amongst organisms, thus suggesting an integral role in gene regulation. *mcg4* may also be involved in some way in tissue-specific or temporal regulation of certain genes, thus making it a potential target for modulating expression of those downstream effectors.

## EXAMPLE 6

Nucleotide sequence data generated from cosmid clone cSRL-72c4 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) was aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul *et al* 1990) and was found to match numerous human and mouse entries (Table 4 and Figure 2). These matching ESTs were further used to identify overlapping entries in the EST database (Table 5). The nucleotide sequences of these human ESTs were compiled using MacVector 4.2.1 software (IBI-Kodak) to produce the cDNA sequence shown in Figure 1. EST entries AA074703 and AA134788 are closely related at the nucleotide level to *mcg4* and it is, therefore, likely that *mcg4* is a member of a newly discovered gene family (Figure 8).

The cDNA sequence of *mcg4* was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al*, 1990) at the National Center for Biotechnology Information (<http://www.ncbi.nih.gov.nlm>). As the protein appeared to be novel, a translation of the longest reading frame for the *mcg4* cDNA was aligned to the EST database using the program TBLASTN, which performed a dynamic translation of the EST database in all 6 frames. The search results indicated that the nematode *C. elegans* had an MCG4-like protein (Figure 3), with the matching domains containing a spatial sequence of Cysteine and Histidine residues which resembled a zinc-finger structure (Figure 4). The program BLASTP was used, therefore, to conduct sensitive searches of the protein databases for similar zinc-finger motifs. A weak match to the putative zinc-finger domain was observed for another protein from *C. elegans* (Figure 5) and a poorer match for the GATA-binding transcription factor from *S. pombe* (Figure 6). The putative initiation codon of human *mcg4* is not preceded by an in-frame stop codon and it is therefore possible that the cDNA described in Figure 1 is a truncated form. However, sequence alignment of human and mouse *mcg4* ESTs showed a lower degree of nucleotide conservation prior to the assigned initiation codon, thus supporting the notion that the region represents the 5' UTR (Figure 9). To determine the expression pattern of *mcg4*, 15µg of the total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% w/v MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer

- 41 -

using 20 x SSC (Sambrook *et al*, 1989). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (<sup>32</sup>P-dCTP) cDNA probe (Church and Gilbert, 1984) for *mcg4*. After washes in 0.1 x SSC/0.1% w/v SDS at 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg4* is expressed as a 1.6kb  
5 message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 7).

#### EXAMPLE 7

A human gene (designated *mcg7*) was identified and isolated from chromosome 11q13 which  
10 encodes a protein that bears striking homology with guanine nucleotide exchange factors (GEFs) from a wide variety of organisms (Fig. 12).

#### EXAMPLE 8

15 The composite *mcg7* cDNA sequence is at least 2.4kb in length and Figure 13(a) shows a predicted translation product of at least 609 amino acids beginning at methionine 120. An alternative start site due to alternate exon splicing (indicated in lower case) may yield a protein of 671 amino acids starting at methionine 58 (Fig. 13a).

20

#### EXAMPLE 9

An *mcg7* homologue from *C. elegans* has been identified, the product of which is highly conserved with that of MCG7 (Fig. 14). There are several salient features of the protein which have been underlined in Fig. 14 - namely: a guanine nucleotide binding region, a diacylglycerol  
25 binding region, and "EF-hand"-calcium binding regions. In addition, there are several potential cAMP, protein kinase C, and casein kinase II phosphorylation sites, as well as a number of potential sites for glycosylation (not indicated).

#### EXAMPLE 10

30

A number of partial human and murine EST clones exist for *mcg7*. The GenBank database

contains a cDNA (Acc. no. Y12336) encoding a full-length open reading frame (ORF) for human *mcg7* as well as a partial murine *mcg7* ORF (Y12339). In addition, the complete genomic sequence of the human *mcg7* gene is contained within GenBank entry AC000134.

5

**EXAMPLE 11**

The best characterised GEFs are members of the family of *ras* oncoproteins, which play a pivotal role in signal transduction and when mutated are responsible for tumour development. A variety of therapeutic regimes for cancer treatment have been designed to specifically interfere with the *ras* signalling pathways. There is potential, therefore that the product of *mcg7* could also be a target for such clinical strategies.

**EXAMPLE 12**

15 The nucleotide sequence for *mcg7* cDNA was extended 5' with genomic DNA sequence from Genbank accession number AC000134 (positions 1-321) and analysed for additional coding sequence 5' to the putative initiation codon (nt 681-683) (Fig. 16). An additional in-frame ATG occurs at position nt 495-497 when the alternatively splice exon (position nt 504-609) is present (also shown in Fig. 13(a)). This closely matches the Kozak consensus. When this exon is  
20 absent, then the ATG is not in-frame and other possible initiation codons are absent (resulting translation shown in lower case lettering) (also shown in Fig. 13(b)). Further evidence that the initiation codon at position nt 681-683 is the true initiation site is given in Figure 15.

Alignment of human and a partial murine *mcg7* cDNA sequences is shown in Figure 15. The  
25 putative initiation codon is at position nt 360-362. Both murine ESTs appear to have an upstream in-frame stop codon at position nt 326-328, downstream of the differentially spliced exon and the sequence alignment thus suggests that this region represents the 5' UTR of *mcg7*.

Furthermore, similarity with the *C. elegans* homologue strongly suggest that the ATG codon at  
30 position nt 360-362 encodes the N-terminus of MCG7.



### EXAMPLE 13

Figure 17 shows data from experiments indicating that a truncated version of MCG7 when expressed as a GST fusion protein (construct B in Fig. 18) can function as a Ras-guanine nucleotide exchange factor. In brief, Ras (unprocessed and as a GST fusion protein) is loaded with <sup>3</sup>H-GDP then incubated in the presence of excess cold GTP ± GST-MCG7. Full details of this assay can be found in Porfiri *et al.*

### EXAMPLE 14

10

Nucleotide sequence data generated from cosmid clone cSRL-20h12 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) were aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul *et al.*, 1990) and was found to match GenBank entries T78563 (clone 113434) TO9103 (clone HIBBP12) and AA035643 (clone 471819). EST clones 113434 and 471819 were obtained from Genome Systems Inc. and these DNAs were sequenced on both strands with gene-specific primers (Table 5) to generate the cDNA sequence of *mcg7* shown in Figures 13(a) and (b).

The cDNA sequence of *mcg7* was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al.*, 1990) and the coding region was assigned on the basis of showing homology to the *C. elegans* protein F25B3.3 (Figure 14). The *mcg7* cDNA composite was suspected to contain a single nucleotide error that originated from clone 471819 and the correct nucleotide sequence was, therefore, sought by reverse transcription-polymerase chain reaction (RT-PCR) of the cDNA fragment from a human cDNA pool. Total RNA was extracted from a human lymphoblastoid cell line using an RNeasy Mini Kit (Qiagen). cDNA synthesis was conducted with the reverse transcriptase Superscript II RNaseH- (GIBCO, BRL) and random hexamers using the procedure recommended by the manufacturer (GIBCO, BRL). One fortieth of the cDNA mix was subjected to 35 cycles of PCR using the following cycling conditions: 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 90 seconds. The 50µl reaction mix consisted of 1x reaction buffer (Dade Scientific), 2mM dNTP mix, 20pmol of primers (see Table 6) MCG7UF (within the

variably spliced exon of Figure 13(b), between nucleotide positions 184-201) and SGCADRV2 (between nucleotide positions 866-846 of Figure 13(a)) and 10 units of Dynazyme (Dade Scientific). The resulting PCR product was cloned into the pGEM-T vector (Promega) using standard methodology and sequenced using gene-specific primers. The correct nucleotide  
 5 sequence of *mcg7* (as shown in Figure 13(a)) matches that of the recently release GenBank entry Y12336. A partial mouse *mcg7* cDNA sequence can also be found in GenBank entry Y12339.

### EXAMPLE 15

10 The coding sequence of *mcg7* was cloned into vectors for expression in both bacterial and mammalian cells. In addition to the full-length constructs, the deletion constructs shown in Figure 18 were designed to retain the guanine nucleotide exchange (GEF) domain. For prokaryotic expression, the *mcg7* coding region was inserted downstream of and in-frame with the Sj26 cassette of the pGEX (Pharmacia) series of vectors (Smith and Johnson, 1988) using  
 15 standard cloning techniques (Sambrook *et al*, 1989). For mammalian expression, the *mcg7* coding sequence was first *myc*-tagged at the N-terminus and then ligated into the expression vector pc Exv-n using standard cloning techniques. Ligation junctions of the constructs were sequences as the cloning strategies inadvertently changed or introduced additional amino acids as shown below.

20

**Construct (A):** EST clone 113434 was digested with *Apa*I (Figure 13(a), nucleotide positions 1022 to >2416 (within the vector)), blunt-ended with T4 DNA polymerase according to the specifications of the manufacturer (New England Biolab) and ligated into the *Sma*I site of pGEX-3X.

25

Sequence of the pGEX and *mcg7* (underlined) junction:

pGEX-3X                      *mcg7* (1022)  
 Sj26 ... GGG ATC CCC CTG GTC [SEQ ID NO:19]

additional amino acids    Gly   Ile   Pro

30

**Construct (B):** EST clone 113434 was digested with *Eco*RI (Figure 13(a), nucleotide

pGEX-1 *mcg7* (695)

additional amino acids    Glu    Phe    Gly    Thr    Ser

15

-----myc-tag-----	vector	BamHI	mcg7 5' UTR (337)	start
ATGGAGCAGAAGCTGATCTCCGAGGAGGACCTG	CCCCGGGCAGCT	ggatcc	CAGCCCACCCCGCGCCGCGGCCATG	
M E Q K L I S E E D L	P G A A G S		A A H P A P A A M	
	-----additional amino acids-----			

25 (Beranger *et al*, 1994).

30 with *Bam*HI and *Sma*I. Digestion with *Bam*HI removed the *myc*-tag of Construct (C).

Sequence of the pGEX and *mcg7* [SEQ ID NO:23/24] (underlined) junction:

- 46 -

pGEX-2      *Bam*HI *mcg7* (337)  
 Sj26      ... gga tcc GCA GCC CAC CCC GCG CCG GCG GCC ATG  
             Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met  
             -----additional amino acids-----

5

**EXAMPLE 16**

Overnight bacterial cultures containing the pGEX plasmid were used to inoculate 500ml of Luria Broth media containing 50µg/ml ampicillin. The cultures were grown to an OD of ~0.8 and then  
 10 induced with 1mM of IPTG for up to 3 hours at 37°C. The bacteria were pelleted and resuspended in 15 ml of STE buffer (10mM Tris pH 8.0, 150 mM NaCl and 1mM EDTA) with 1 mg/ml lysozyme. The mixture was left on ice for more than 1 hour and subsequent steps were performed at 4°C. Protease inhibitors aprotinin, pepstatin and leupeptin were added at final concentrations of 25µg/ml, prior to the addition of Triton-X-100 (2% v/v final) and n-lauroyl  
 15 sarcosine (1.5% w/v final). The lysate was sonicated for ~1 minute and pelleted at 14,000 x g for 15 minutes. 100 µl of 50% w/v glutathione-sephadex bead slurry (in PBS) was added per ml of supernatant. Following a 30 minute incubation at 4°C, the beads were washed three times with NETN (20mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5% NP40), once with NETN-HS (equivalent to NETN but with 1M NaCl), and once in NETN. The bound protein  
 20 was directly analysed by SDS-polyacrylamide gel electrophoresis (PAGE) as described below or the bound protein was eluted from the beads with the following elution buffer (50mM Tris pH 8.0, 150mM NaCl, 5mM MgCl<sub>2</sub>, 1mM DTT, 10mM reduced glutathione) for use in GDP release assays.

25

**EXAMPLE 17**

Twenty microlitres of GST-sepharose-bound MCG7 were added to an equal volume of 2 x  
 30 sample loading dye (100mM Tris pH6.8, 2% v/v mercaptoethanol, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol), boiled for 5 min and loaded onto a 7.5% w/v SDS-PAGE gel (Sambrook *et al*, 1989). The Coomassie brilliant blue stained gel (Sambrook *et al*, 1989)

typically displayed a protein doublet, running between 87-95 kDa consisting of the MCG7-GST fusion and a slightly smaller, co-purified contaminating *E. coli* protein of ~105kDa. The calculated molecular weight of full-length MCG7 is 77.5 kDa (Construct (D)) and the GST component has a molecular weight of 26kDa, hence, the recombinant protein runs slightly smaller than predicted. A Western blot of the same gel probed with anti-GST antibody yields an MCG7-specific band at the same position as that of the stained gel.

### EXAMPLE 18

10 Assumptions: (a) GST-Ras molecular weight = 50 kD; (b) Concentration of GST-Ras solution = 1mg/ml = 20 $\mu$ M; (c) [ $^3$ H]-GDP is 1mCi/ml and 13.3Ci/mmol, therefore [H]-GDP concentration = 75  $\mu$ M and 1pmol [ $^3$ H]-GDP=15,466 cpm; (d) Elution buffer = Buffer E = 20 mM Tris-Cl, pH7.5; 50mM NaCl; 5mM MgCl<sub>2</sub>; 1mM DTT (added just before use). Buffer E + BSA= Buffer E+1mg/ml BSA (added just before use).

15

Mix together, in the following order and mix well after each addition:

10 $\mu$ l (=10 $\mu$ g) GST-Ras (@1mg/ml in Buffer E), 463 $\mu$ l Buffer E + BSA, 7 $\mu$ l [ $^3$ H]-GDP, 10ml 490  $\mu$ M EDTA. Incubate @ RT for 10 min. Add 10 $\mu$ l 0.5 M MgCl<sub>2</sub> and mix well. Incubate @ RT for 10 min. Place on ice. During the first incubation the excess EDTA concentration is 5mM, during the second incubation the excess Mg concentration is 5mM. The [ $^3$ H]-GDP concentration is 1 $\mu$ M and the final concentration of GST-Ras is 400nM. Thus 20ml of the final mix will contain 8pmol of GST-Ras protein. Specific activity of GDP is 15,446 cpm/pmol x (1/1.4) = 11,047 cpm/pmol.

25

### EXAMPLE 19

Exchange Ras with labelled GDP as above. Add unlabelled GTP (stock = 100mM, pH7) to 1 mM. Adjust Mg concentration by adding 5 $\mu$ l 0.5 EDTA to labelled Ras, 5 $\mu$ l 0.5M EDTA to 500 $\mu$ l MCG7, and 5 $\mu$ l 0.5M EDTA to 500 $\mu$ l Buffer E + BSA. On ice set up microfuge tubes with 40 $\mu$ l Ras-GDP (in triplicate) with 40 $\mu$ l MCG7 or Buffer E + BSA (control). Transfer tubes to heat block @ 25°C and incubate for 10, 20 or 30 min. Stop exchange reactions with 1ml of

- 48 -

ice cold buffer E and place on ice. Pre-soak nitrocellulose filters, pore size  $45\mu\text{m}$ , in Buffer E. Assemble the vacuum manifold apparatus (Millipore) with wet filters and plug the wells with rubber bunds. Switch on the vacuum pump. Remove the first plug, aliquot the sample and once it has been sucked through, wash the filter with 10ml of ice cold Buffer E. Remove next plug  
5 etc and continue round the manifold. Take manifold apart. Pin the filters to a pin board reserved for [ $^3\text{H}$ ]. Air dry. Take up in 4ml scintillation fluid and count. These studies have been carried out with a truncated MCG7-GST fusion protein (amino acids 341 of Figure 13a to stop encoded within construct B).

10

**EXAMPLE 20**

A human gene was identified from chromosome 11q13 that encodes a new member of the DnaJ family of proteins (designated MCG18). This gene (*mcg18*) is expressed as an ~1.4kb mRNA (Fig. 28) and is predicted to encode a 241 amino acid product (Fig. 19).

15

**EXAMPLE 21**

MCG18 has partial homology to *E. coli* dnaJ and other human DnaJ family members in that it contains the J domain (Fig. 20).

20

**EXAMPLE 22**

MCG18 has greatest homology to functionally undefined proteins from *C. elegans* (Fig. 21) and *S. pombe* (Fig. 22) that also feature the J domain but maintain sequence similarity through the  
25 central and C-terminal regions of the proteins.

**EXAMPLE 23**

The J domain is proposed to mediate interaction with heat shock protein (Hsp70) 70 and consist  
30 of some 70 amino acids, frequently located at the N-terminus of the protein. One of these proteins, tumorous imaginal discs (Tid58) from *Drosophila virilis* (Fig. 23) functions as a

tumour suppressor.

#### EXAMPLE 24

- 5 A comparison of homology between MCG18 and human DnaJ proteins HDJ-2/H5DJ, HDJ-1/HSP40 and HSJ1 is shown in Fig. 24.

#### EXAMPLE 25

- 10 During the sequence characterisation of the *VRF/VEGFB* promoter region on cosmid CLGW4 [Grimmond *et al*, 1996], which maps to chromosome 11q13 the inventors identified a sequence that exactly matched numerous human and mouse expressed sequence tags (ESTs) in the EST database from a gene which we designated *mcg18*. EST clones for human (GenBank accession number T69741, clone 108172; accession number H40901, clone 177008) and mouse *mcg18*  
15 (accession number W34884, clone 350966; accession number W64183, clone 385535) were obtained from Genome Systems Inc. and sequenced with the gene-specific primers shown in Table 7. The EST clones listed in Table 8 were also utilised in generating the full-length coding sequence for human (Figure 19) and mouse (Figure 25) *mcg18*. The EST database also contained *mcg18* cDNA entries that were alternately (or partially) spliced, and in order to  
20 understand their ability to encode new polypeptides, the gene structure of *mcg18* was determined by sequencing human and mouse genomic templates with gene-specific primers.

Genomic fragments containing the human [Grimmond *et al*, 1996] and murine genes [Townson *et al*, 1996] have been previously reported. Cosmid CLGW4 contains the entire human gene  
25 and  $\lambda$ 121 contains the entire mouse gene, as determined by direct sequencing of the templates with the oligonucleotides listed in Table 7. Plasmids containing sub-fragments of  $\lambda$ 121 and cosmid CLGW4 were prepared using plasmid purification kits (Qiagen) and sequenced as described previously [Grimmond *et al*, 1996; Townson *et al*, 1996] using primers designed against cDNA and genomic sequences. The BLAST suite of programs [Altschul *et al*, 1990]  
30 was used to compare the sequence data against the nucleotide and protein databases at the National Center for Biotechnology Information (<http://www.ncbi.nih.gov.nlm>). The sequence

data were compiled using MacVector 4.2.1 software (IBI-Kodak). ClustalW sequence alignments [Thompson *et al*, 1994] were conducted using the Australian National Genome Information Service computer faculty at the University of Sydney, Australia.

- 5 The cDNA sequence of human *mcg18* (Figure 19) was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX [Altschul *et al*, 1990] and the coding region was identified on the basis of showing homology to the DnaJ family of proteins (Figure 20). The DnaJ domain is encoded within the longest open reading frame and the assigned initiation codon is preceded by an in-frame stop codon (Figure 10 27). Similar database search results were obtained for the mouse *mcg18* cDNA, and the alignment of human and mouse protein sequences is shown in Figure 26. MCG18 has greatest homology to gene products from *C. elegans* (Figure 21) and *S. pombe* (Figure 22). Although it shares a similar J-domain, MCG18 does not contain other domains described for the tumour suppressor gene from *D. virilis* (Figure 23), nor is it a homologue of other reported human J- 15 domain-containing proteins (Figure 24).

To determine the expression pattern of *mcg18*, 15µg of total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer 20 using 20 x SSC (Sambrook *et al*, 1986). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (<sup>32</sup>P-dCTP) cDNA probe (Church and Gilbert, 1984) for *mcg18*. After washes in 0.1 x SSC/0.1% w/v SDS for 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg18* is expressed as a 1.4kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 28).



- 51 -

TABLE 4

ESTs matching *mcg4*

accession number	seq. run	organism	score	E value	N
gb AA399110 AA399110	zt89e06.s1	Soares testis NHT Homo sa...	1136	4.0e-168	2
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone 2...	1521	5.3e-168	4
gb AA514406 AA514406	nf57d01.s1	NCI_CGAP_Co3 Homo sapiens...	931	5.5e-166	3
gb AA544946 AA544946	vk38e02.r1	Soares mouse mammary glan...	1207	8.4e-164	2
gb AA450076 AA450076	zx42a04.s1	Soares total fetus Nb2HF8...	691	2.3e-160	4
gb AA535731 AA535731	nf88f07.s1	NCI_CGAP_Co3 Homo sapiens...	796	3.5e-158	4
gb W79710 W79710	zd86f01.r1	Soares fetal heart NbHH19...	1644	1.1e-157	4
gb AA503531 AA503531	ne47e08.s1	NCI_CGAP_Co3 Homo sapiens...	736	4.0e-156	4
gb AA450132 AA450132	zx42a04.r1	Soares total fetus Nb2HF8...	1955	3.9e-155	1
gb AA398068 AA398068	zt89f06.r1	Soares testis NHT Homo sa...	1315	5.4e-148	2
gb W60405 W60405	zd29h08.r1	Soares fetal heart NbHH19...	1022	1.8e-139	4
gb W81382 W81382	zd86f01.s1	Soares fetal heart NbHH19...	605	3.5e-125	5
gb AA047617 AA047617	zf13f07.s1	Soares fetal heart NbHH19...	922	4.6e-125	2
gb AA282175 AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapien...	1577	2.0e-123	1
gb AA242159 AA242159	my30d04.r1	Barstead mouse pooled org...	866	7.7e-117	2
gb AA068680 AA068680	mm61a05.r1	Stratagene mouse embryoni...	1280	1.6e-98	1
gb W46766 W46766	zc36b07.s1	Soares senescent fibrobla...	506	9.6e-92	3
gb N93704 N93704	zb51c04.s1	Soares fetal lung NbHL19W...	584	9.0e-91	4
gb AA155210 AA155210	mr98e01.r1	Stratagene mouse embryoni...	840	7.6e-87	2
gb AA366022 AA366022	EST76915	Pineal gland II Homo sapien...	1077	2.4e-81	1
gb AA037691 AA037691	zk34h12.s1	Soares pregnant uterus Nb...	949	2.1e-80	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor ...	1016	3.1e-76	1
dbj C00696 C00696	HUMGS0008251	Human Gene Signature, ...	1009	1.2e-75	1
gb T98249 T98249	ye59a07.s1	Homo sapiens cDNA clone 1...	998	6.7e-75	1
gb W21588 W21588	zb51c04.r1	Soares fetal lung NbHL19W...	484	1.1e-69	4
gb H32171 H32171	EST107015	Rattus sp. cDNA 5' end.	828	1.1e-60	1
gb AA108092 AA108092	mm89e06.r1	Stratagene mouse embryoni...	782	1.3e-60	2
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4Nb...	665	2.5e-60	2
gb AA037690 AA037690	zk34h12.r1	Soares pregnant uterus Nb...	540	9.4e-53	2
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapien...	535	5.4e-48	2
gb N46760 N46760	yy51g06.r1	Homo sapiens cDNA clone 2...	665	9.5e-47	1
gb W23584 W23584	zc71d03.s1	Soares fetal heart NbHH19...	457	1.8e-44	2
gb W42214 W42214	mc69h09.r1	Soares mouse embryo NbME1...	460	1.3e-38	3
gb AA244877 AA244877	mx25a04.r1	Soares mouse NML Mus musc...	429	2.9e-25	1
gb W32939 W32939	zc07h03.r1	Soares parathyroid tumor ...	320	4.8e-18	1

- 52 -

TABLE 5

ESTs matching AA074703 (*mcg4*-related cDNA)

Database: Non-redundant Database of GenBank EST Division  
 1,222,625 sequences; 449,352,662 total letters.

			Smallest Sum		
Sequences producing High-scoring Segment Pairs:			High	Probability	
accession number	seq. run	organism	Score	P(N)	N
			score	E value	N
gb AA074703 AA074703	zm76g07.r1	Stratagene neuroepitheli...	2071	4.0e-167	1
gb AA068680 AA068680	mm61a05.r1	Stratagene mouse embryon...	1270	4.4e-145	4
gb AA134788 AA134788	zm81g02.r1	Stratagene neuroepitheli...	946	1.3e-144	5
gb AA399110 AA399110	zt89e06.s1	Soares testis NHT Homo s...	520	8.7e-119	6
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone ...	582	9.6e-110	7
gb AA282175 AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapie...	771	9.4e-80	3
gb W81382 W81382	zd86f01.s1	Soares fetal heart NbHH1...	329	1.6e-75	6
gb AA544946 AA544946	vk38e02.r1	Soares mouse mammary gla...	644	9.6e-63	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor...	294	4.5e-42	4
gb W57106 W57106	md57c12.r1	Soares mouse embryo NbME...	394	1.9e-30	2
gb AA244877 AA244877	mx25a04.r1	Soares mouse NML Mus mus...	162	2.1e-27	4
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4N...	230	3.7e-23	3
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapie...	139	2.3e-19	3
gb H32171 H32171	EST107015	Rattus sp. cDNA 5' end.	207	2.6e-10	2
gb W79710 W79710	zd86f01.r1	Soares fetal heart NbHH1...	157	0.0073	1

**TABLE 6**  
*mcg7*-specific oligonucleotides

5	name	sequence (5' to 3')	SEQ ID NOs.
	M1044R	GGA CAA AGT GTG TGA TGA ACC	SEQ ID NO:25
	MCG7-GEF-REV2	CTC ATC CTC CGT CTG ATA CTG	SEQ ID NO:26
	M7R	GTA GAT GTG GAT CAG CTT GG	SEQ ID NO:27
	MCG7 CA FOR	AGG TGG AGA ATG GTC AAGG	SEQ ID NO:28
10	MCG7-GEF-REV	GTC ATA GTC TGT CTC CTA CT	SEQ ID NO:29
	MCG7 GEF FOR	ACA TAG ACA GCG TGC CTA CC	SEQ ID NO:30
	MCG7-PKC-REV	TAC AAC CTT AGG GAC ACC AG	SEQ ID NO:31
	MCG7-PKC-FOR	TGC TGA GCC TGC TCA CGG TG	SEQ ID NO:32
	T09103F	CAA GTG AAC AGC ACG TCC	SEQ ID NO:33
15	M7F	GAC TAT CTC AAG GAC CAG CTG	SEQ ID NO:34
	MCG7UF	GGT TCG GTC CGA GCC CGG	SEQ ID NO:35
	SGCADRV2	GGA GCG ATA CTC CAA GTA GGT	SEQ ID NO:36

**TABLE 7**  
***mcg18*-SPECIFIC OLIGONUCLEOTIDES**

	<b>name</b>	<b>sequence 5' to 3'</b>
5	HVESTF	AGC GGG CCA GGC CCC TTC [SEQ ID NO:37]
	HV195F	CAT CCT GGT CCA ATG CGC TC [SEQ ID NO:38]
	HV387F2	GCA CTG AGG AAG TTA AAC GAG C [SEQ ID NO:39]
	HV408R	GCT CGT TTA ACT TCC TCA GTG C [SEQ ID NO:40]
	EXON1REV	GCT CAG CTC CAC AAA GCG GCT [SEQ ID NO:41]
10	HVEST426F	ACC AGC TCC GCT CAG GTA G [SEQ ID NO:42]
	HVEST623R	TCC AGG AGC TGT GTG TTT GG [SEQ ID NO:43]
	SGVESTF3	CCA GTT TCA CAG CGT GAG G [SEQ ID NO:44]
	HVEST631R	CAG CAT GAG GAG GAG GCA G [SEQ ID NO:45]

**TABLE 8**  
**EST CLONE SEQUENCES USED TO GENERATE HUMAN AND MOUSE**  
***mcg18* cDNA SEQUENCE COMPOSITES**

<u>EST clone number</u>	<u>organism</u>	<u>GenBank accession number</u>
1g2815	human	D45683
001-T2-18	human	F17225
273748	human	N37043
177008	human	H40901 and H40939
258011	human	N30776
276887	human	N44004
108172	human	T69741
307529	human	W21083 and W32579
342027	human	W60283
354288	mouse	W44038
350966	mouse	W348844
426261	mouse	AA002868
368185	mouse	W53911
385535	mouse	W64183
404472	mouse	W82959
406437	mouse	W83482

**BIBLIOGRAPHY**

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *J. Mol. Biol.* 215: 403-410.
2. Church, G., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* 18: 1991-1995.
3. Porfiri *et al.* *J. Biol. Chem.* 269: 22672-22677 (1994).
4. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, USA.
5. Smith, D.B., and Johnson, K.S. (1988) *Gene* 67: 31-40.
6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) *Nucleic Acids Res.* 22: 4673-4680.
7. Beranger, F., Paterson, H., Powers, S., de Gunzburg, J. and Hancock, J.F. (1994) *Molecular and Cellular Biology* 14: 744-758.
8. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G (1996) *Genome Res.* 6: 124-131.
9. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun.* 220: 922-928.

- 57 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: (OTHER THAN US): The Council of The Queensland Institute of Medical Research  
(US ONLY): HAYWARD Nicholas, SILINS Ginters, GRIMMOND Sean, GARTSIDE Michael and HANCOCK, John

(ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 45

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DAVIES COLLISON CAVE  
(B) STREET: 1 LITTLE COLLINS STREET  
(C) CITY: MELBOURNE  
(D) STATE: VICTORIA  
(E) COUNTRY: AUSTRALIA  
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: PCT INTERNATIONAL  
(B) FILING DATE: 22-MAY-1998  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6973  
(B) FILING DATE: 23-MAY-1997  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6974  
(B) FILING DATE: 23-MAY-1997  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6972  
(B) FILING DATE: 23-MAY-1997

- 58 -

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PP1459

(B) FILING DATE: 22-JAN-1998

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PP1460

(B) FILING DATE: 22-JAN-1998

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PP1458

(B) FILING DATE: 22-JAN-1998

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: HUGHES, DR E JOHN L

(C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: +61 3 9254 2777

(B) TELEFAX: +61 3 9254 2770

(C) TELEX: AA 31787



- 59 -

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Cys Xaa Xaa Cys Xaa Gly Xaa Gly  
 5

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1242 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 30..959

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCAGTAAACA CAGAGACTGG GGATCGATC	ATG GGG CTT TGT AAG TGC CCC AAG	53
	Met Gly Leu Cys Lys Cys Pro Lys	
	1 5	
AGA AAG GTG ACC AAC CTG TTC TGC TTC GAA CAT CGG GTC AAC GTC TGC	101	
Arg Lys Val Thr Asn Leu Phe Cys Phe Glu His Arg Val Asn Val Cys		
10 15 20		
GAG CAC TGC CTG GTA GCC AAT CAC GCC AAG TGC ATC GTC CAG TCC TAC	149	
Glu His Cys Leu Val Ala Asn His Ala Lys Cys Ile Val Gln Ser Tyr		
25 30 35 40		
CTG CAA TGG CTC CAA GAT AGC GAC TAC AAC CCC AAT TGC CGC CTG TGC	197	
Leu Gln Trp Leu Gln Asp Ser Asp Tyr Asn Pro Asn Cys Arg Leu Cys		
45 50 55		
AAC ATA CCC CTG GCC AGC CGA GAG ACG ACC CGC CTT GTC TGC TAT GAT	245	
Asn Ile Pro Leu Ala Ser Arg Glu Thr Thr Arg Leu Val Cys Tyr Asp		
60 65 70		
CTC TTT CAC TGG GCC TGC CTC AAT GAA CGT GCT GCC CAG CTA CCC CGA	293	
Leu Phe His Trp Ala Cys Leu Asn Glu Arg Ala Ala Gln Leu Pro Arg		
75 80 85		
AAC ACG GCA CCT GCC GGC TAT CAG TGC CCC AGC TGC AAT GGC CCC ATC	341	
Asn Thr Ala Pro Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ile		
90 95 100		
TTC CCC CCA ACC AAC CTG GCT GGC CCC GTG GCC TCC GCA CTG AGA GAG	389	
Phe Pro Pro Thr Asn Leu Ala Gly Pro Val Ala Ser Ala Leu Arg Glu		
105 110 115 120		

- 60 -

AAG CTG GCC ACA GTC AAC TGG GCC CGG GCA GGA CTG GGC CTC CCT CTG	437
Lys Leu Ala Thr Val Asn Trp Ala Arg Ala Gly Leu Gly Leu Pro Leu	
125 130 135	
ATC GAT GAG GTG GTG AGC CCA GAG CCC GAG CCC CTC AAC ACG TCT GAC	485
Ile Asp Glu Val Val Ser Pro Glu Pro Glu Pro Leu Asn Thr Ser Asp	
140 145 150	
TTC TCT GAC TGG TCT AGT TTT AAT GCC AGC AGT ACC CCT GGA CCA GAG	533
Phe Ser Asp Trp Ser Ser Phe Asn Ala Ser Ser Thr Pro Gly Pro Glu	
155 160 165	
GAG GTA GAC AGC GCC TCT GCT GCC CCA GCC TTC TAC AGC CGA GCC CCC	581
Glu Val Asp Ser Ala Ser Ala Ala Pro Ala Phe Tyr Ser Arg Ala Pro	
170 175 180	
CGG CCC CCA GCT TCC CCA GGC CGG CCC GAG CAG CAC ACA GTG ATC CAC	629
Arg Pro Pro Ala Ser Pro Gly Arg Pro Glu Gln His Thr Val Ile His	
185 190 195 200	
ATG GGC AAT CCT GAG CCC TTG ACT CAC GCC CCT AGG AAG GTG TAT GAT	677
Met Gly Asn Pro Glu Pro Leu Thr His Ala Pro Arg Lys Val Tyr Asp	
205 210 215	
ACG CGG GAT GAT GAC CGG ACA CCA GGC CTC CAT GGA GAC TGT GAC GAT	725
Thr Arg Asp Asp Asp Arg Thr Pro Gly Leu His Gly Asp Cys Asp Asp	
220 225 230	
GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTG CTA	773
Asp Lys Tyr Arg Arg Arg Pro Ala Leu Gly Trp Leu Ala Arg Leu Leu	
235 240 245	
AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG	821
Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg	
250 255 260	
GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT	869
Ala Gly Leu Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu	
265 270 275 280	
GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC	917
Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn	
285 290 295	
CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA	962
Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser *	
300 305 310	
GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT	1022
AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA	1082
CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG	1142
GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC	1202
ATGAAGGAGC TGGCAGGTGG AAATAAACAA CAACTTTATT	1242

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 310 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- 61 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Gly Leu Cys Lys Cys Pro Lys Arg Lys Val Thr Asn Leu Phe Cys
 1           5           10           15
Phe Glu His Arg Val Asn Val Cys Glu His Cys Leu Val Ala Asn His
          20           25           30
Ala Lys Cys Ile Val Gln Ser Tyr Leu Gln Trp Leu Gln Asp Ser Asp
          35           40           45
Tyr Asn Pro Asn Cys Arg Leu Cys Asn Ile Pro Leu Ala Ser Arg Glu
          50           55           60
Thr Thr Arg Leu Val Cys Tyr Asp Leu Phe His Trp Ala Cys Leu Asn
          65           70           75           80
Glu Arg Ala Ala Gln Leu Pro Arg Asn Thr Ala Pro Ala Gly Tyr Gln
          85           90           95
Cys Pro Ser Cys Asn Gly Pro Ile Phe Pro Pro Thr Asn Leu Ala Gly
          100          105          110
Pro Val Ala Ser Ala Leu Arg Glu Lys Leu Ala Thr Val Asn Trp Ala
          115          120          125
Arg Ala Gly Leu Gly Leu Pro Leu Ile Asp Glu Val Val Ser Pro Glu
          130          135          140
Pro Glu Pro Leu Asn Thr Ser Asp Phe Ser Asp Trp Ser Ser Phe Asn
          145          150          155          160
Ala Ser Ser Thr Pro Gly Pro Glu Glu Val Asp Ser Ala Ser Ala Ala
          165          170          175
Pro Ala Phe Tyr Ser Arg Ala Pro Arg Pro Pro Ala Ser Pro Gly Arg
          180          185          190
Pro Glu Gln His Thr Val Ile His Met Gly Asn Pro Glu Pro Leu Thr
          195          200          205
His Ala Pro Arg Lys Val Tyr Asp Thr Arg Asp Asp Asp Arg Thr Pro
          210          215          220
Gly Leu His Gly Asp Cys Asp Asp Asp Lys Tyr Arg Arg Arg Pro Ala
          225          230          235          240
Leu Gly Trp Leu Ala Arg Leu Leu Arg Ser Arg Ala Gly Ser Arg Lys
          245          250          255
Arg Pro Leu Thr Leu Leu Gln Arg Ala Gly Leu Leu Leu Leu Leu Gly
          260          265          270
Leu Leu Gly Phe Leu Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg
          275          280          285
Ala Ala Ala Asp Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His
          290          295          300
Ile Arg Val Gly Pro Ser
          305          310

```

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- 62 -

- (A) LENGTH: 2415 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..2188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC TCC CCA AAA TAT TCC	47
Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser	
1 5 10 15	
CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG CTG TCC	95
His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser	
20 25 30	
CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG CCT GCA GGA	143
Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly	
35 40 45	
AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA ACG CAG CGC CTG	191
Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu	
50 55 60	
TGT GGC CGC GGG ACT CAA GGC TGG CCT GGC TCA AGT GAA CAG CAC GTC	239
Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val	
65 70 75	
CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT TCT GGG GTG GAC GAG CTG	287
Gln Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu	
80 85 90 95	
GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC	335
Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly	
100 105 110	
CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC CTG GAC	383
Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp	
115 120 125	
AAG GGC TGC ACG GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC	431
Lys Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe	
130 135 140	
GAT GAC TCC GGG AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC	479
Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu	
145 150 155	
ATG ATG CAC CCC TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG	527
Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu	
160 165 170 175	
CTC CAC ATC TAC CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG	575
Leu His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln	
180 185 190	
GTG AAA ACG TGC CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG	623
Val Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala	
195 200 205	

- 63 -

GAG TTT GAC TTG AAC CCG GAG TTG GCT GAG CAG ATC AAG GAG CTG AAG Glu Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys 210 215 220	671
GCT CTG CTA GAC CAA GAA GGG AAC CGA CGG CAC AGC AGC CTA ATC GAC Ala Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp 225 230 235	719
ATA GAC AGC GTC CCT ACC TAC AAG TGG AAG CGG CAG GTG ACT CAG CGG Ile Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg 240 245 250 255	767
AAC CCT GTG GGA CAG AAA AAG CGC AAG ATG TCC CTG TTG TTT GAC CAC Asn Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His 260 265 270	815
CTG GAG CCC ATG GAG CTG GCG GAG CAT CTC ACC TAC TTG GAG TAT CGC Leu Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg 275 280 285	863
TCC TTC TGC AAG ATC CTG TTT CAG GAC TAT CAC AGT TTC GTG ACT CAT Ser Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His 290 295 300	911
GGC TGC ACT GTG GAC AAC CCC GTC CTG GAG CGG TTC ATC TCC CTC TTC Gly Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe 305 310 315	959
AAC AGC GTC TCA CAG TGG GTG CAG CTC ATG ATC CTC AGC AAA CCC ACA Asn Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr 320 325 330 335	1007
GCC CCG CAG CGG GCC CTG GTC ATC ACA CAC TTT GTC CAC GTG GCG GAG Ala Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu 340 345 350	1055
AAG CTG CTA CAG CTG CAG AAC TTC AAC ACG CTG ATG GCA GTG GTC GGG Lys Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly 355 360 365	1103
GGC CTG AGC CAC AGC TCC ATC TCC CGC CTC AAG GAG ACC CAC AGC CAC Gly Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His 370 375 380	1151
GTT AGC CCT GAG ACC ATC AAG CTC TGG GAG GGT CTC ACG GAA CTA GTG Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val 385 390 395	1199
ACG GCG ACA GGC AAC TAT GGC AAC TAC CGG CGT CGG CTG GCA GCC TGT Thr Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys 400 405 410 415	1247
GTG GGC TTC CGC TTC CCG ATC CTG GGT GTG CAC CTC AAG GAC CTG GTG Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val 420 425 430	1295
GCC CTG CAG CTG GCA CTG CCT GAC TGG CTG GAC CCA GCC CGG ACC CGG Ala Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg 435 440 445	1343
CTC AAC GGG GCC AAG ATG AAG CAG CTC TTT AGC ATC CTG GAG GAG CTG Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu 450 455 460	1391
GCC ATG GTG ACC AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu 465 470 475	1439

- 64 -

CTG AGC CTG CTC ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu 480 485 490 495	1487
CTG TAC CAG CTG TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro 500 505 510	1535
ACC AGC CCC ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu 515 520 525	1583
GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val 530 535 540	1631
GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val 545 550 555	1679
GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly 560 565 570 575	1727
AAC TTC CCT TAC CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp 580 585 590	1775
GGC TGC ATC AGC AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser 595 600 605	1823
TCT GTG TTG GGG GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser 610 615 620	1871
AAC TCC TTG CGC CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG Asn Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu 625 630 635	1919
GGC ATC TAC AAG CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys 640 645 650 655	1967
CAC AAG CAG TGC AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala 660 665 670	2015
CAG AGT GTG AGC CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His 675 680 685	2063
AGC CAC CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG Ser His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg 690 695 700	2111
CGA GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val 705 710 715	2159
GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG TGTTGGATC Glu Asp Gly Val Phe Asp Ile His Leu 720 725	2208
AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA GCAGGGAGCC TGGGGGTGTC	2268
GGGGCAGGAG GCTGGGGATG GGGGTGGGAT ATGAGGGTGG CATGCAGCTG AGGGCAGGGC	2328

- 65 -

CAGGGCTGGT GTCCCTAAGG TTGTACAGAC TCTTGTGAAT ATTTGTATTT TCCAGATGGA 2388  
 ATAAAAAGGC CCGTGTAATT AACCTTC 2415

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 728 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser His  
 1 5 10 15  
 Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser Pro  
 20 25 30  
 Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly Arg  
 35 40 45  
 Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu Cys  
 50 55 60  
 Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val Gln  
 65 70 75 80  
 Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu Gly  
 85 90 95  
 Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro  
 100 105 110  
 Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys  
 115 120 125  
 Gly Cys Thr Val Glu Glu Leu Arg Gly Cys Ile Glu Ala Phe Asp  
 130 135 140  
 Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met  
 145 150 155 160  
 Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu  
 165 170 175  
 His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val  
 180 185 190  
 Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu  
 195 200 205  
 Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala  
 210 215 220  
 Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile  
 225 230 235 240  
 Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn  
 245 250 255  
 Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu  
 260 265 270

- 66 -

Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser  
                   275                                  280                                  285  
 Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His Gly  
                   290                                  295                                  300  
 Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe Asn  
   305                                  310                                  315                                  320  
 Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr Ala  
                                   325                                  330                                  335  
 Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu Lys  
                                   340                                  345                                  350  
 Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly Gly  
                                   355                                  360                                  365  
 Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His Val  
                                   370                                  375                                  380  
 Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr  
   385                                  390                                  395                                  400  
 Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val  
                                   405                                  410                                  415  
 Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala  
                                   420                                  425                                  430  
 Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu  
                                   435                                  440                                  445  
 Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala  
                                   450                                  455                                  460  
 Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu  
   465                                  470                                  475                                  480  
 Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu  
                                   485                                  490                                  495  
 Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr  
                                   500                                  505                                  510  
 Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu  
                                   515                                  520                                  525  
 Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu  
                                   530                                  535                                  540  
 His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp  
   545                                  550                                  555                                  560  
 Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn  
                                   565                                  570                                  575  
 Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly  
                                   580                                  585                                  590  
 Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser  
                                   595                                  600                                  605  
 Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn  
                                   610                                  615                                  620  
 Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly



- 67 -

625		630		635		640
Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His	645		650		655	
Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln	660		665		670	
Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser	675		680		685	
His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg	690		695		700	
Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu	705		710		715	720
Asp Gly Val Phe Asp Ile His Leu	725					

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2309 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 254..2083

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG	60
CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC	120
TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAGC CCCATGGGAA	180
CGGGGTTCGG TCCGAGCCCG GTGGGAGGCT CCCGGAGCGC AGCCTGGGCC CAGCCCACCC	240
CGCGCCGGCG GCC ATG GCA GGC ACC CTG GAC CTG GAC AAG GGC TGC ACG	289
Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr	
1 5 10	
GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC GAT GAC TCC GGG	337
Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly	
15 20 25	
AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC ATG ATG CAC CCC	385
Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro	
30 35 40	
TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG CTC CAC ATC TAC	433
Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr	
45 50 55 60	
CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG GTG AAA ACG TGC	481
Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys	
65 70 75	
CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG GAG TTT GAC TTG	529

- 68 -

His	Leu	Val	Arg	Tyr	Trp	Ile	Ser	Ala	Phe	Pro	Ala	Glu	Phe	Asp	Leu	
			80					85					90			
AAC	CCG	GAG	TTG	GCT	GAG	CAG	ATC	AAG	GAG	CTG	AAG	GCT	CTG	CTA	GAC	577
Asn	Pro	Glu	Leu	Ala	Glu	Gln	Ile	Lys	Glu	Leu	Lys	Ala	Leu	Leu	Asp	
		95					100					105				
CAA	GAA	GGG	AAC	CGA	CGG	CAC	AGC	AGC	CTA	ATC	GAC	ATA	GAC	AGC	GTC	625
Gln	Glu	Gly	Asn	Arg	Arg	His	Ser	Ser	Leu	Ile	Asp	Ile	Asp	Ser	Val	
	110					115					120					
CCT	ACC	TAC	AAG	TGG	AAG	CGG	CAG	GTG	ACT	CAG	CGG	AAC	CCT	GTG	GGA	673
Pro	Thr	Tyr	Lys	Trp	Lys	Arg	Gln	Val	Thr	Gln	Arg	Asn	Pro	Val	Gly	
125					130					135					140	
CAG	AAA	AAG	CGC	AAG	ATG	TCC	CTG	TTG	TTT	GAC	CAC	CTG	GAG	CCC	ATG	721
Gln	Lys	Lys	Arg	Lys	Met	Ser	Leu	Leu	Phe	Asp	His	Leu	Glu	Pro	Met	
				145					150					155		
GAG	CTG	GCG	GAG	CAT	CTC	ACC	TAC	TTG	GAG	TAT	CGC	TCC	TTC	TGC	AAG	769
Glu	Leu	Ala	Glu	His	Leu	Thr	Tyr	Leu	Glu	Tyr	Arg	Ser	Phe	Cys	Lys	
			160					165					170			
ATC	CTG	TTT	CAG	GAC	TAT	CAC	AGT	TTC	GTG	ACT	CAT	GGC	TGC	ACT	GTG	817
Ile	Leu	Phe	Gln	Asp	Tyr	His	Ser	Phe	Val	Thr	His	Gly	Cys	Thr	Val	
		175					180					185				
GAC	AAC	CCC	GTC	CTG	GAG	CGG	TTC	ATC	TCC	CTC	TTC	AAC	AGC	GTC	TCA	865
Asp	Asn	Pro	Val	Leu	Glu	Arg	Phe	Ile	Ser	Leu	Phe	Asn	Ser	Val	Ser	
	190					195					200					
CAG	TGG	GTG	CAG	CTC	ATG	ATC	CTC	AGC	AAA	CCC	ACA	GCC	CCG	CAG	CGG	913
Gln	Trp	Val	Gln	Leu	Met	Ile	Leu	Ser	Lys	Pro	Thr	Ala	Pro	Gln	Arg	
205					210					215					220	
GCC	CTG	GTC	ATC	ACA	CAC	TTT	GTC	CAC	GTG	GCG	GAG	AAG	CTG	CTA	CAG	961
Ala	Leu	Val	Ile	Thr	His	Phe	Val	His	Val	Ala	Glu	Lys	Leu	Leu	Gln	
				225					230					235		
CTG	CAG	AAC	TTC	AAC	ACG	CTG	ATG	GCA	GTG	GTC	GGG	GGC	CTG	AGC	CAC	1009
Leu	Gln	Asn	Phe	Asn	Thr	Leu	Met	Ala	Val	Val	Gly	Gly	Leu	Ser	His	
			240					245					250			
AGC	TCC	ATC	TCC	CGC	CTC	AAG	GAG	ACC	CAC	AGC	CAC	GTT	AGC	CCT	GAG	1057
Ser	Ser	Ile	Ser	Arg	Leu	Lys	Glu	Thr	His	Ser	His	Val	Ser	Pro	Glu	
		255					260					265				
ACC	ATC	AAG	CTC	TGG	GAG	GGT	CTC	ACG	GAA	CTA	GTG	ACG	GCG	ACA	GGC	1105
Thr	Ile	Lys	Leu	Trp	Glu	Gly	Leu	Thr	Glu	Leu	Val	Thr	Ala	Thr	Gly	
	270					275					280					
AAC	TAT	GGC	AAC	TAC	CGG	CGT	CGG	CTG	GCA	GCC	TGT	GTG	GGC	TTC	CGC	1153
Asn	Tyr	Gly	Asn	Tyr	Arg	Arg	Arg	Leu	Ala	Ala	Cys	Val	Gly	Phe	Arg	
285					290					295					300	
TTC	CCG	ATC	CTG	GGT	GTG	CAC	CTC	AAG	GAC	CTG	GTG	GCC	CTG	CAG	CTG	1201
Phe	Pro	Ile	Leu	Gly	Val	His	Leu	Lys	Asp	Leu	Val	Ala	Leu	Gln	Leu	
				305					310					315		
GCA	CTG	CCT	GAC	TGG	CTG	GAC	CCA	GCC	CGG	ACC	CGG	CTC	AAC	GGG	GCC	1249
Ala	Leu	Pro	Asp	Trp	Leu	Asp	Pro	Ala	Arg	Thr	Arg	Leu	Asn	Gly	Ala	
			320					325					330			
AAG	ATG	AAG	CAG	CTC	TTT	AGC	ATC	CTG	GAG	GAG	CTG	GCC	ATG	GTG	ACC	1297
Lys	Met	Lys	Gln	Leu	Phe	Ser	Ile	Leu	Glu	Glu	Leu	Ala	Met	Val	Thr	
		335					340					345				

AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG CTG AGC CTG CTC Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu 350 355 360	1345
ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG CTG TAC CAG CTG Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu 365 370 375 380	1393
TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA ACC AGC CCC ACG Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr 385 390 395	1441
AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG GAG TGG ACC TCG Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser 400 405 410	1489
GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG GAG CAC ATC GAG Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu 415 420 425	1537
AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC GAT GGG GAT GGC Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly 430 435 440	1585
CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG AAC TTC CCT TAC His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr 445 450 455 460	1633
CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT GGC TGC ATC AGC Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly Cys Ile Ser 465 470 475	1681
AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC TCT GTG TTG GGG Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly 480 485 490	1729
GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC AAC TCC TTG CGC Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg 495 500 505	1777
CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG GGC ATC TAC AAG Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys 510 515 520	1825
CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC CAC AAG CAG TGC Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys 525 530 535 540	1873
AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC CAG AGT GTG AGC Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser 545 550 555	1921
CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC AGC CAC CAT CAC Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser His His 560 565 570	1969
CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG CGA GGC TCC AGG Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Gly Ser Arg 575 580 585	2017
CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG GAG GAT GGG GTG Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val 590 595 600	2065
TTT GAC ATC CAC TTG TAATAGATGC TGTGGTTGGA TCAAGGACTC ATTCCTGCCT Phe Asp Ile His Leu 605 610	2120

- 70 -

TGGAGAAAAT ACTTCAACCA GAGCAGGGAG CCTGGGGGTG TCGGGGCAGG AGGCTGGGGA 2180  
 TGGGGGTGGG ATATGAGGGT GGCATGCAGC TGAGGGCAGG GCCAGGGCTG GTGTCCCTAA 2240  
 GGTGTACAG ACTCTTGTGA ATATTGTAT TTTCCAGATG GAATAAAAAG GCCCGTGTA 2300  
 TTAACCTTC 2309

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 609 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
 1 5 10 15  
 Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly Lys Val Arg Asp  
 20 25 30  
 Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro Trp Tyr Ile Pro  
 35 40 45  
 Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr Gln Gln Ser Arg  
 50 55 60  
 Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys His Leu Val Arg  
 65 70 75 80  
 Tyr Trp Ile Ser Ala Phe Pro Ala Glu Phe Asp Leu Asn Pro Glu Leu  
 85 90 95  
 Ala Glu Gln Ile Lys Glu Leu Lys Ala Leu Leu Asp Gln Glu Gly Asn  
 100 105 110  
 Arg Arg His Ser Ser Leu Ile Asp Ile Asp Ser Val Pro Thr Tyr Lys  
 115 120 125  
 Trp Lys Arg Gln Val Thr Gln Arg Asn Pro Val Gly Gln Lys Lys Arg  
 130 135 140  
 Lys Met Ser Leu Leu Phe Asp His Leu Glu Pro Met Glu Leu Ala Glu  
 145 150 155 160  
 His Leu Thr Tyr Leu Glu Tyr Arg Ser Phe Cys Lys Ile Leu Phe Gln  
 165 170 175  
 Asp Tyr His Ser Phe Val Thr His Gly Cys Thr Val Asp Asn Pro Val  
 180 185 190  
 Leu Glu Arg Phe Ile Ser Leu Phe Asn Ser Val Ser Gln Trp Val Gln  
 195 200 205  
 Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg Ala Leu Val Ile  
 210 215 220  
 Thr His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe  
 225 230 235 240  
 Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser  
 245 250 255

- 71 -

Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu Thr Ile Lys Leu  
 260 265 270  
 Trp Glu Gly Leu Thr Glu Leu Val Thr Ala Thr Gly Asn Tyr Gly Asn  
 275 280 285  
 Tyr Arg Arg Arg Leu Ala Ala Cys Val Gly Phe Arg Phe Pro Ile Leu  
 290 295 300  
 Gly Val His Leu Lys Asp Leu Val Ala Leu Gln Leu Ala Leu Pro Asp  
 305 310 315 320  
 Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala Lys Met Lys Gln  
 325 330 335  
 Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr Ser Leu Arg Pro  
 340 345 350  
 Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu Thr Val Ser Leu  
 355 360 365  
 Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu Ser Leu Gln Arg  
 370 375 380  
 Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr Ser Cys Thr Pro  
 385 390 395 400  
 Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser Ala Ala Lys Pro  
 405 410 415  
 Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu Lys Met Val Glu  
 420 425 430  
 Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly His Ile Ser Gln  
 435 440 445  
 Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr Leu Ser Ala Phe  
 450 455 460  
 Gly Asp Leu Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met  
 465 470 475 480  
 Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly Gly Arg Met Gly  
 485 490 495  
 Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys  
 500 505 510  
 Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys  
 515 520 525  
 Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu  
 530 535 540  
 Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser Leu Glu Gly Ser  
 545 550 555 560  
 Ala Pro Ser Pro Ser Pro Met His Ser His His His Arg Ala Phe Ser  
 565 570 575  
 Phe Ser Leu Pro Arg Pro Gly Arg Arg Gly Ser Arg Pro Pro Glu Ile  
 580 585 590  
 Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val Phe Asp Ile His  
 595 600 605  
 Leu

- 72 -

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 832 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 11..733

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCCGCGCGCC	ATG	CCG	CCC	TTA	CTG	CCC	CTG	CGC	CTG	TGC	CGG	CTG	TGG	49
	Met	Pro	Pro	Leu	Leu	Pro	Leu	Arg	Leu	Cys	Arg	Leu	Trp	
	1				5					10				
CCC CGC AAC CCT CCC TCC CGG CTC CTC GGA GCG GCC GCC GGG CAG CGG	97													
Pro Arg Asn Pro Pro Ser Arg Leu Leu Gly Ala Ala Ala Gly Gln Arg														
15 20 25														
TCC AGA CCC AGT ACT TAT TAT GAA CTG TTG GGG GTG CAT CCT GGT GCC	145													
Ser Arg Pro Ser Thr Tyr Tyr Glu Leu Leu Gly Val His Pro Gly Ala														
30 35 40 45														
AGC ACT GAG GAA GTT AAA CGA GCT TTC TTC TCC AAG TCC AAA GAG CTG	193													
Ser Thr Glu Glu Val Lys Arg Ala Phe Phe Ser Lys Ser Lys Glu Leu														
50 55 60														
CAC CCA GAC CGG GAC CCT GGG AAC CCA AGC CTG CAC AGC CGC TTT GTG	241													
His Pro Asp Arg Asp Pro Gly Asn Pro Ser Leu His Ser Arg Phe Val														
65 70 75														
GAG CTG AGC GAG GCA TAC CGT GTG CTC AGC CGT GAG CAG AGC CGC CGC	289													
Glu Leu Ser Glu Ala Tyr Arg Val Leu Ser Arg Glu Gln Ser Arg Arg														
80 85 90														
AGC TAT GAT GAC CAG CTC CGC TCA GGT AGT CCC CCA AAG TCT CCA CGA	337													
Ser Tyr Asp Asp Gln Leu Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg														
95 100 105														
ACC ACA GTC CAT GAC AAG TCT GCC CAC CAA ACA CAC AGC TCC TGG ACA	385													
Thr Thr Val His Asp Lys Ser Ala His Gln Thr His Ser Ser Trp Thr														
110 115 120 125														
CCC CCC AAC GCA CAG TAC TGG TCC CAG TTT CAC AGC GTG AGG CCA CAG	433													
Pro Pro Asn Ala Gln Tyr Trp Ser Gln Phe His Ser Val Arg Pro Gln														
130 135 140														
GGG CCC CAG TTG AGG CAG CAG CAA CAC AAA CAA AAC AAA CAA GTG CTG	481													
Gly Pro Gln Leu Arg Gln Gln Gln His Lys Gln Asn Lys Gln Val Leu														
145 150 155														
GGG TAC TGC CTC CTC CTC ATG CTG GCG GGC ATG GGC CTG CAC TAC ATT	529													
Gly Tyr Cys Leu Leu Leu Met Leu Ala Gly Met Gly Leu His Tyr Ile														
160 165 170														
GCC TTC AGG AAG GTG AAG CAG ATG CAC CTT AAC TTC ATG GAT GAA AAG	577													
Ala Phe Arg Lys Val Lys Gln Met His Leu Asn Phe Met Asp Glu Lys														
175 180 185														

- 73 -

GAT CGG ATC ATC ACA GCC TTC TAC AAC GAA GCC CGG GCA CGG GCC AGG 625  
 Asp Arg Ile Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg  
 190 195 200 205

GCC AAC AGA GGC ATC CTT CAG CAG GAG CGA CAA CGG CTA GGG CAG CGG 673  
 Ala Asn Arg Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg  
 210 215 220

CAG CCG CCA CCA TCC GAG CCA ACC CAA GGC CCC GAG ATC GTG CCC CGG 721  
 Gln Pro Pro Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg  
 225 230 235

GGC GCC GGC CCC TGA GGGGCTC ACCTGGATGG GGCCTGCAGT GCGTTCCCGC 773  
 Gly Ala Gly Pro \*

TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTGCGAG 832

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 241 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Pro Pro Leu Leu Pro Leu Arg Leu Cys Arg Leu Trp Pro Arg Asn  
 1 5 10 15

Pro Pro Ser Arg Leu Leu Gly Ala Ala Ala Gly Gln Arg Ser Arg Pro  
 20 25 30

Ser Thr Tyr Tyr Glu Leu Leu Gly Val His Pro Gly Ala Ser Thr Glu  
 35 40 45

Glu Val Lys Arg Ala Phe Phe Ser Lys Ser Lys Glu Leu His Pro Asp  
 50 55 60

Arg Asp Pro Gly Asn Pro Ser Leu His Ser Arg Phe Val Glu Leu Ser  
 65 70 75 80

Glu Ala Tyr Arg Val Leu Ser Arg Glu Gln Ser Arg Arg Ser Tyr Asp  
 85 90 95

Asp Gln Leu Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg Thr Thr Val  
 100 105 110

His Asp Lys Ser Ala His Gln Thr His Ser Ser Trp Thr Pro Pro Asn  
 115 120 125

Ala Gln Tyr Trp Ser Gln Phe His Ser Val Arg Pro Gln Gly Pro Gln  
 130 135 140

Leu Arg Gln Gln Gln His Lys Gln Asn Lys Gln Val Leu Gly Tyr Cys  
 145 150 155 160

Leu Leu Leu Met Leu Ala Gly Met Gly Leu His Tyr Ile Ala Phe Arg  
 165 170 175

Lys Val Lys Gln Met His Leu Asn Phe Met Asp Glu Lys Asp Arg Ile  
 180 185 190

Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg Ala Asn Arg

- 74 -

195	200	205
Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg Gln Pro Pro		
210	215	220
Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg Gly Ala Gly		
225	230	235
240		
Pro		

SEQ ID Nos: 10-18 25-36

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 300 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 170..300

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGATTTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG	60
CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC	120
TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAG CCC CAT	175
Pro His	
1	
GGG AAC GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC	223
Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser	
5 10 15	
CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC	271
Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp	
20 25 30	
CTG GAC AAG GGC TGC ACG GTG GAG GAG CT	300
Leu Asp Lys Gly Cys Thr Val Glu Glu Leu	
35 40	

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein



- 75 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Pro His Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu  
 1 5 10 15  
 Arg Ser Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr  
 20 25 30  
 Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
 35 40

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGATCCCCC TGGTC

15

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asp Val Asp Glu Glu Asp Glu Val Glu Asp Ile Glu Phe  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids

- 76 -

(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asp His Asp Arg Asp Gly Phe Ile Ser Gln Glu Glu Phe  
1 5 10

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met  
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Val Asp Met Asp Gly Gln Ile Ser Lys Asp Glu Leu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe Asn  
1 5 10 15

Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser Arg  
20 25 30

- 77 -

Leu Lys Glu Thr His  
35

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Phe Val His Val Ala Lys His Leu Arg Lys Ile Asn Asn Phe Asn  
1 5 10 15  
 Thr Leu Met Ser Val Val Gly Gly Ile Thr His Ser Ser Val Ala Arg  
20 25 30  
 Leu Ala Lys Thr Tyr  
35

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys Arg His  
1 5 10 15  
 Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg  
20 25 30  
 Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val  
35 40 45  
 Glu Cys  
50

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

His Asn Phe His Glu Thr Thr Phe Leu Thr Pro Thr Thr Cys Asn His

- 78 -

1	5	10	15
Cys Asn Lys	Leu Leu Trp Gly Ile	Leu Arg Gln Gly Phe	Lys Cys Lys
	20	25	30
Asp Cys Gly	Leu Ala Val His Ser	Cys Cys Lys Ser	Asn Ala Val Ala
	35	40	45
Glu Cys			
50			

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGGATCCCCC TGCTC

15

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAATTCGGCA CGAGCCGACG G

21

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 78 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGGAGCAGA AGCTGATCTC CGAGGAGGAC CTGCCCCGGGG CAGCTGGATC CGCAGCCCAC

60

CCCGCGCCGG CGGCCATG

78

## (2) INFORMATION FOR SEQ ID NO:22:

- 79 -

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Pro Gly Ala Ala Gly  
1 5 10 15  
Ser Ala Ala His Pro Ala Pro Ala Ala Met  
20 25

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGATCCGCAG CCCACCCCGC GCCGGCGGCC ATG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met  
5 10

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 80 -

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:  
GGACAAAGTG TGTGATGAAC C 21
- (2) INFORMATION FOR SEQ ID NO:26:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:  
CTCATCCTCC GTCTGATACT G 21
- (2) INFORMATION FOR SEQ ID NO:27:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:  
GTAGATGTGG ATCAGCTTGG 20
- (2) INFORMATION FOR SEQ ID NO:28:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:  
AGGTGGAGAA TGGTCAAGG 19
- (2) INFORMATION FOR SEQ ID NO:29:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:  
GTCATAGTCT GTCTCCTACT 20

- 81 -

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ACATAGACAG CGTGCCTACC

20

## (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TACAACCTTA GGGACACCAG

20

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TGCTGAGCCT GCTCACGGTG

20

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CAAGTGAACA GCACGTCC

18

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:

- 82 -

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GACTATCTCA AGGACCAGCT G

21

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGTTCGGTCC GAGCCCGG

18

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGAGCGATAC TCCAAGTAGG T

21

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AGCGGGCCAG GCCCCTTC

18

(2) INFORMATION FOR SEQ ID NO:38:



- 83 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATCCTGGTC CAATGCGCTC

20

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCACTGAGGA AGTTAAACGA GC

22

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCTCGTTTAA CTCCTCAGT GC

22

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCTCAGCTCC ACAAAGCGGC T

21

(2) INFORMATION FOR SEQ ID NO:42:

- 84 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ACCAGCTCCG CTCAGGTAG

19

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TCCAGGAGCT GTGTGTTTGG

20

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CCAGTTTCAC AGCGTGAGG

19

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CAGCATGAGG AGGAGGCAG

19

## CLAIMS:

1. An isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
2. An isolated nucleic acid molecule according to claim 1 wherein the regulator comprises a zinc finger domain of an  $(\text{HC}_3)_2$  type.
3. An isolated nucleic acid molecule according to claim 2 wherein the sequence of nucleotides or complementary sequence of nucleotides is selected from:
  - (i) a nucleotide sequence set forth in SEQ ID NO:2;
  - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
  - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
  - (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
4. An isolated nucleic acid molecule according to claim 1 wherein said gene regulator is a guanine nucleotide exchange factor (GEF) or a derivative thereof.
5. An isolated nucleic acid molecule according to claim 4 wherein the sequence of nucleotides is selected from:
  - (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
  - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
  - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
  - (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

6. An isolated nucleic acid molecule according to claim 1, wherein said gene regulator is a heat shock protein or is a heat shock binding protein or a derivative thereof.

7. An isolated nucleic acid molecule according to claim 6, wherein the sequence of nucleotides is selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

8. A genetic construct comprising a vector portion and a gene portion comprising a regulator of gene expression or a derivative thereof .

9. A genetic construct according to claim 8 wherein the gene portion comprises a zinc finger domain of  $(HC_3)_2$  type.

10. A genetic construct according to claim 9 wherein the gene portion comprises a nucleotide sequence selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

11. A genetic construct according to claim 8 wherein said gene portion is a nucleotide exchange factor (GEF) or derivative thereof.

12. A genetic construct according to claim 11 wherein the gene portion comprises a nucleotide sequence selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

13. A genetic construct according to claim 8 wherein the gene portion is a heat shock protein or a derivative thereof or a heat shock binding protein or derivative thereof.

14. A genetic construct according to claim 13 wherein the gene portion comprises a nucleotide sequence selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

15. A nucleic acid molecule encoding a gene regulator having the identifying characteristics of a molecule selected from MCG4, MCG7 and MCG18 having respective amino acid sequences of SEQ ID NO:3, SEQ ID NO: 5 or 7 and SEQ ID NO:9.

16. A method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

17. A method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

18. A method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

19. A method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

20. A method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

21. A method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

22. A method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

23. A method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

24. A method for detecting MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

## FIGURE 1

TCAGTAAACA CAGAGACTGG GGATCGATC ATG GGG CTT TGT AAG TGC CCC AAG	53
Met Gly Leu Cys Lys Cys Pro Lys	
1 5	
AGA AAG GTG ACC AAC CTG TTC TGC TTC GAA CAT CGG GTC AAC GTC TGC	101
Arg Lys Val Thr Asn Leu Phe Cys Phe Glu His Arg Val Asn Val Cys	
10 15 20	
GAG CAC TGC CTG GTA GCC AAT CAC GCC AAG TGC ATC GTC CAG TCC TAC	149
Glu His Cys Leu Val Ala Asn His Ala Lys Cys Ile Val Gln Ser Tyr	
25 30 35 40	
CTG CAA TGG CTC CAA GAT AGC GAC TAC AAC CCC AAT TGC CGC CTG TGC	197
Leu Gln Trp Leu Gln Asp Ser Asp Tyr Asn Pro Asn Cys Arg Leu Cys	
45 50 55	
AAC ATA CCC CTG GCC AGC CGA GAG ACG ACC CGC CTT GTC TGC TAT GAT	245
Asn Ile Pro Leu Ala Ser Arg Glu Thr Thr Arg Leu Val Cys Tyr Asp	
60 65 70	
CTC TTT CAC TGG GCC TGC CTC AAT GAA CGT GCT GCC CAG CTA CCC CGA	293
Leu Phe His Trp Ala Cys Leu Asn Glu Arg Ala Ala Gln Leu Pro Arg	
75 80 85	
AAC ACG GCA CCT GCC GGC TAT CAG TGC CCC AGC TGC AAT GGC CCC ATC	341
Asn Thr Ala Pro Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ile	
90 95 100	
TTC CCC CCA ACC AAC CTG GCT GGC CCC GTG GCC TCC GCA CTG AGA GAG	389
Phe Pro Pro Thr Asn Leu Ala Gly Pro Val Ala Ser Ala Leu Arg Glu	
105 110 115 120	
AAG CTG GCC ACA GTC AAC TGG GCC CGG GCA GGA CTG GGC CTC CCT CTG	437
Lys Leu Ala Thr Val Asn Trp Ala Arg Ala Gly Leu Gly Leu Pro Leu	
125 130 135	
ATC GAT GAG GTG GTG AGC CCA GAG CCC GAG CCC CTC AAC ACG TCT GAC	485
Ile Asp Glu Val Val Ser Pro Glu Pro Glu Pro Leu Asn Thr Ser Asp	
140 145 150	
TTC TCT GAC TGG TCT AGT TTT AAT GCC AGC AGT ACC CCT GGA CCA GAG	533
Phe Ser Asp Trp Ser Ser Phe Asn Ala Ser Ser Thr Pro Gly Pro Glu	
155 160 165	
GAG GTA GAC AGC GCC TCT GCT GCC CCA GCC TTC TAC AGC CGA GCC CCC	581
Glu Val Asp Ser Ala Ser Ala Ala Pro Ala Phe Tyr Ser Arg Ala Pro	
170 175 180	
CGG CCC CCA GCT TCC CCA GGC CGG CCC GAG CAG CAC ACA GTG ATC CAC	629
Arg Pro Pro Ala Ser Pro Gly Arg Pro Glu Gln His Thr Val Ile His	
185 190 195 200	
ATG GGC AAT CCT GAG CCC TTG ACT CAC GCC CCT AGG AAG GTG TAT GAT	677
Met Gly Asn Pro Glu Pro Leu Thr His Ala Pro Arg Lys Val Tyr Asp	
205 210 215	



ACG CGG GAT GAT GAC CGG ACA CCA GGC CTC CAT GGA GAC T. GAC GAT 775  
 Thr Arg Asp Asp Asp Arg Thr Pro Gly Leu His Gly Asp Cys Asp Asp  
 220 225 230

GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTG CTA 773  
 Asp Lys Tyr Arg Arg Arg Pro Ala Leu Gly Trp Leu Ala Arg Leu Leu  
 235 240 245

AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG 821  
 Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg  
 250 255 260

GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT 869  
 Ala Gly Leu Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu  
 265 270 275 280

GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC 917  
 Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn  
 285 290 295

CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962  
 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser  
 300 305 310

GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGT 1022

AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082

CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142

GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202

ATGAAGGAGC TGGCAGGTGG AAATAAACAA CAACTTTATT 1242

3/32

**Figure 2**

gb|AA155210|AA155210 mr98e01.r1 Stratagene mouse embryonic carcinoma  
 (#937317) Mus musculus cDNA clone 605496 5'

Query: 1 MGLCKCPKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQSDYNPNCRLCNIP 60  
 MGLCKCPKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQSDYNPNCRLCN PL  
 Sbjct: 98 MGLCKCPKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQSDYNPNCRLCNTP 277

**Figure 3**

dbj|D75913|CELK111G3F C.elegans cDNA clone yk111g3 : 5' end, single read.

Query: 7 PKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQSDYNPNCRLCNIPASRETT 66  
 PKRKVTNLF +EHRVNVCE LV NH C+VQSYL WL D DY+PNC LC L +T  
 Sbjct: 1 PKRKVTNLFXYEHRVNVCELVNDHPNCVVQSYLTWLTDDQYDPNCSLCRTTLXEGDTI 180

Query: 67 RLVCYDLFWACLNERRAAQLPRNTAPAGYQCP 98 98 PSCNGPIFFPPNQ 109  
 RL C L HW C +E P TAP GY+CP P C+ +FPP+Q  
 Sbjct: 181 RLNCLHLLHWKCFDEWXGNFPDITAPXGYRCP 276 275 PCCSQEVFPFPDQ 310

Figure 4

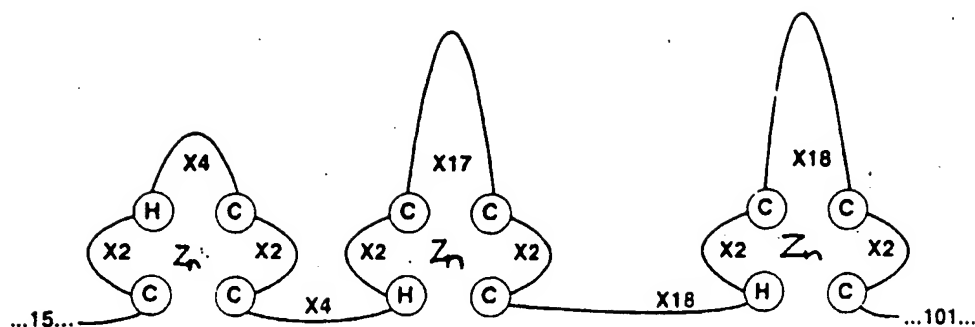


Figure 5

sp|P46580|YLB5\_CAEEL HYPOTHETICAL 146.8 KD PROTEIN C34E10.5 IN  
CHROMOSOME III gi|500728 (U10402) C34E10.5 gene product  
[Caenorhabditis elegans]

Query: 56 CNIPLASRETTTLVCYDLFWACLNERAAQLPRNTAPAGYQCPSC 100  
C+I L ++ + L C LF W C+ E A + + + +CP C  
Sbjct: 1222 CSICLENKNPSALFCGHLFCWTCIQEHAVAATSSASTSSARCPQC 1266

Figure 6

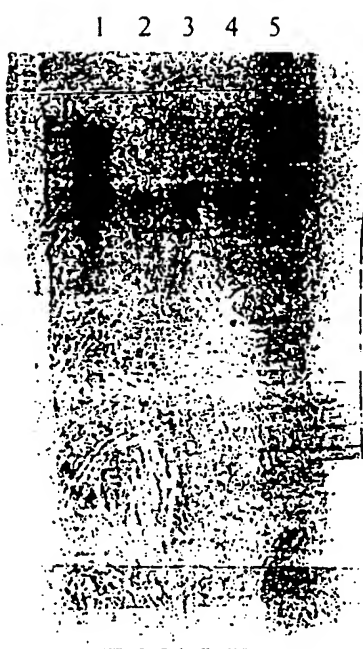
gi|703468 (L29051) homologous to GATA-binding transcription factor  
[Schizosaccharomyces pombe]

Query: 35 CIVQSYLQWLQSDYNPNCRLCNI 58  
C + +W +D NP C C +  
Sbjct: 175 CATTNTPKWRRDESGNPICNACGL 198

Query: 162 SSTPGPEEVDASAAAPAFYSQAPRPPASGRPEQHTVIHMGNEPLTHAPRKVYDTRDDO 221  
+S PEE S S S P+ SP + +Q +I P +V + D  
Sbjct: 441 ASLLNPEEPPSNSDKQPSMNGPKSEVSPSQSQAPLIQSSTSPVSLQFPPEVQGSNVDK 500

Query: 222 RTPGLH 227  
R L+  
Sbjct: 501 RNYALN 506

Figure 7



Score = 818 (226.0 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103  
Identities = 206/259 (79%), Positives = 206/259 (79%), Strand = Plus / Plus

Score = 230 (63.6 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103  
Identities = 50/55 (90%), Positives = 50/55 (90%), Strand = Plus / Plus

Score = 175 (48.4 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103  
Identities = 39/44 (88%), Positives = 39/44 (88%), Strand = Plus / Plus

Score = 139 (38.4 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103  
Identities = 31/35 (88%), Positives = 31/35 (88%), Strand = Plus / Plus

Score = 133 (36.8 bits), Expect = 6.1e-103, Sum P(5) = 5.1e-103  
Identities = 29/32 (90%); Positives = 29/32 (90%); Strand = Plus / Plus

```

Query:   701  CGGGATGATGACCGGACACCAGGCCTCCATGG  732
          |||||
Sbjct:   305  CGGGATGATGACCGGACAGCAGGCATTCATGG  336

```

## Figure 8 continued

gb|AA134788|AA134788 zm8lg02.r1 Stratagene neuroepithelium (#937231)  
Homo sapiens cDNA clone 532082 5'  
Length = 368

## Plus Strand HSPs:

Score = 563 (155.6 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87  
Identities = 147/190 (77%), Positives = 147/190 (77%), Strand = Plus / Plus

```

Query:   498 CGTCTGACTTCTCTGACTGGTCTAGTTTAAATGCCAGCAGTACCCCTGGACCAGAGGAGG 557
          ||| ||||| ||||| ||| ||||| ||||| ||| ||||| ||||| ||| |||||
Sbjct:   103 CCTCAGACTTCTCTGATTTGGTCCAGCTTTAATGCCACCACCACCTCTGTGCAAGAGGAGA 162

Query:   558 TAGACAGCGCCTCTGCTGCCCCAGCCTTCTACAGCCAGGCCCGCCCGGCCCCAGCTTCCC 617
          || ||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:   163 GAGCCAGCACTCCATCTGCGCCTGCTTTCTATAGCCAGGCTCCCGCCCTCCTCCCTCCC 222

Query:   618 CAGGCCCGGCCCGAGCAGCACACAGTGATCCACATGGGCAA'TCCTGAGCCCTTGACTCAGG 677
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:   223 CAAGCCGTCCCGAGCAGCACACAGTCATACACATGGGGAGTACTGAAGCCCTGGCACACG 282

Query:   678 CCCCTAGGAA 687
          ||||| |||||
Sbjct:   283 CCCCAGGAA 292

```

Score = 454 (125.4 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87  
Identities = 94/98 (95%), Positives = 94/98 (95%), Strand = Plus / Plus

```

Query:   398 GCACTGAGAGAGAAGCTGGCCACAGTCAACTGGGCCCGGGCAGGACTGGGCCTCCCTCTG 457
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:    2 GCACTGAGAGACAAGCTAGCCACAGTCAACTGGGCCCGGGCAGGACTGGGCCTCCCTCTG 61

Query:   458 ATCGATGAGGTGGTGAGCCAGAGCCCGAGCCCTCAA 495
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:   62 ATCGATGAGGTGATAAGCCAGAGCCCGAGCCCTCAA 99

```

Score = 219 (60.5 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87  
Identities = 51/60 (85%), Positives = 51/60 (85%), Strand = Plus / Plus

```

Query:   702 GGGATGATGACCGGACACCAGGCCTCCATGGAGACTGTGACGATGACAAGTACCGACGTC 761
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:   309 GGATTGATGACCGGACAGCAGGCATTCA'TGGAGACTGTGATGATGACAAATACCGCCGCC 368

```

## Figure 9

W32939 human TACCGCCCTTCGGAAACAGTGCAGCGGCCGATCAGTAAACACAGAGACTGGGGATCGATCATGGGGCTTTGTAAAG  
AA242159 mouse CTTCCGCGCTTTTCATTACCGTACGCACCCGCTCA-CGATCGGCATCGCGAGGATCGGTCAATGGGACTTTGCAAG

FIGURE 10

MCG4 MGLCKCPKRR VTNLFCFEHR VNVCEHCLVA NHAKCIVQSY LQWLQSDSYN PNCRLCNIPL 60  
 MCG4 ASRETTTRLVC YDLFWACLN ERAAQLPRNT APAGYQCPSC NGPIFPPTNL AGPVASALRE 120  
 3. [ 229 ] \_\_\_\_\_> \*\*\*x>  
 5. [ 74 ] \_\_\_\_\_> \*\*\*\*>

130 140 150 160 170 180  
 MCG4 KLATVNWARA GLGLPLIDEV VSPEPEPLNT SDFSWSSEFN ASSTPGPEEV DSASAAPAFY  
 1. [ 372 ] \_\_\_\_\_ 20 30 40 50 60  
 2. [ 243 ] \_\_\_\_\_ 30 40 50 60  
 \_\_\_\_\_ aqs\*s\*sip \*\*\*\*\* \*tt\*svq\*\*r a\*tps\*\*\*\*\*>  
 \_\_\_\_\_ \*p

3. [ 229 ] \_\_\_\_\_ 10 20 30 40 50 60  
 \_\_\_\_\_ i\*\*\*\*\*s xrl\*lvql\* chhhlcarge sqh\*icac\*l>

5. [ 74 ] \_\_\_\_\_ 10 20 30 40 50 60  
 \_\_\_\_\_ smr\*\*a q\*\*s\*-sipq tslig-pal- mppp\*lcrr ep\*hlxlli>

190 200 210 220 230 240  
 MCG4 SDAPRPPASP GRPEQHTVIH MGNPEPLTHA PRKVDYTRDD DRTPLHGDC DDKYRRRPA  
 1. [ 372 ] \_\_\_\_\_ 70 80 90 100 110 120  
 \_\_\_\_\_ i\*\*\*\*\*p\*\* s\*\*\*\*\* \*\*st\*a\*a\*\* \*\*\*\*\*pgp \*srhswetvn mtnt-aagl\*>  
 2. [ 243 ] \_\_\_\_\_ 70 80 90  
 \_\_\_\_\_ i\*\*\*\*\*p\*\* s\*\*\*\*\* \*\*st\*a\*a\*\* \*\*\*\*\*>

3. [ 229 ] \_\_\_\_\_ 70 80 90 100 110 120  
 \_\_\_\_\_ gsp\*sslpk\* s\*a-a\*sht\* gey\*s\*g\*r- \*kek\*m\*hg\* \*\*\*a\*i\*\*\*\*\*>  
 4. [ 86 ] \_\_\_\_\_ 70 80 90 100 110 120  
 \_\_\_\_\_ p\*sslpk\* s\*a-a\*sht\* gey\*s\*g\*rp kesi\*h\*gm tgqqaftm\*\*\*\*\*c>

5. [ 74 ] \_\_\_\_\_ 70 80 90 100 110  
 \_\_\_\_\_ arl\*allppq av\*sstqsy w\*vlk\*w-\*t \*qgk\*m\*\*\*\*\* \*\*a\*i\*\*>

6. [ 38 ] \_\_\_\_\_ 100  
 \_\_\_\_\_ \*t \*q\*\*\*\*\*>

250 260 270 280 290 300  
 MCG4 LGWLARLLRS RAGSRKRPLT LLQAGLLLL LGLLGFLALL ALMSRLGRAA ADSDPNLDPL  
 1. [ 372 ] \_\_\_\_\_ 130  
 \_\_\_\_\_ \*\*\*\*\*q\*\*\*\*\*>  
 4. [ 86 ] s\*-->

310  
 MCG4 MNPHIRVGPS

Figure 10 (Continued)

Search Analysis for Sequence: MCG4

Matrix: pam250 matrix

Search from 1 to 310

Score Region from 1 to 310

Date: September 22, 1997

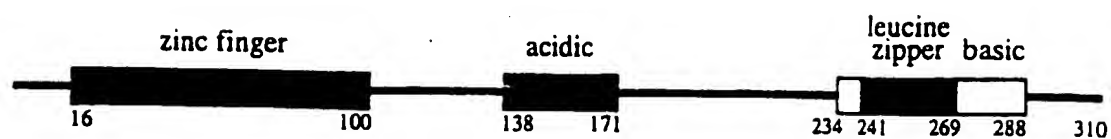
Maximum possible score: 1598

Aligned sequences:

1. = EST AA074703 phase 1 translation
2. = EST AA134788 phase 3 translation
3. = EST AA134788 phase 2 translation
4. = EST AA074703 phase 3 translation
5. = EST AA074703 phase 2 translation
6. = EST AA134788 phase 1 translation



FIGURE 11 Domains of MCG4



zinc finger consensus:  $CX_2HX_4CX_2CX_4HX_2CX_{17}CX_2CX_{18}HX_2CX_{18}CX_2C$

acidic domain consensus: 9/34 negatively charged amino acids, 0/34 positively charged

basic domain consensus: 13/55 positively charged amino acids, 0/55 negatively charged

leucine zipper domain consensus:  $LX_6LX_6RX_6LX_6L$

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa 261)  $LX_6LXLX_6LXLX_6L$  (aa 286)

## FIGURE 12

Sequences producing High-scoring Segment Pairs:			High Score	Smallest Sum Probability P(N)	N
gnl PID e236178	(Z70752) F25B3.3 (Caenorhabditis ele...		307	3.0e-124	8
gi 1293099	(U53884) aimless RasGEF (Dictyosteli...		202	7.8e-22	5
gi 1655941	(U67326) Ras-GRF2 (Mus musculus)		152	3.6e-16	4
pir  S30356	CDC25 protein homolog - yeast (Candi...		150	2.2e-15	3
sp P43069 CC25_CANAL	CELL DIVISION CONTROL PROTEIN 25		150	2.2e-15	3
sp P28818 GNRP_RAT	GUANINE NUCLEOTIDE RELEASING PROTEIN...		166	2.6e-15	3
prf  1814463A	guanine nucleotide-releasing factor ...		166	2.6e-15	3
pir  B46199	nucleotide-exchange-factor homolog c...		167	1.1e-14	1
gnl PID e238680	(X97560) hypothetical protein L1309 ...		158	3.0e-14	3
pir  S22693	CDC25 protein homolog - mouse /gi 50...		167	3.7e-14	2
sp P14771 SC25_YEAST	SCD25 PROTEIN /gi 457494 (M26647) SD...		158	4.6e-14	3
sp P26674 STE6_SCHPO	STE6 PROTEIN /pir  S28098 ste6 prote...		160	5.2e-14	2
pir  S28407	CDC25 protein homolog - mouse		167	1.2e-13	3
sp P27671 GNRP_MOUSE	GUANINE NUCLEOTIDE RELEASING PROTEIN...		167	1.2e-13	3
gi 386047	(S62035) Ras-specific guanine nucleo...		153	2.0e-13	2
sp Q02342 CC25_SACKL	CELL DIVISION CONTROL PROTEIN 25 /pi...		142	4.5e-13	2
pir  S14177	SCD25 protein - yeast (Saccharomyces...		152	5.7e-13	3
gi 433720	(L26584) CDC25 [Homo sapiens]		153	6.0e-13	3
gnl PID e241744	(Z68880) T14G10.2 (Caenorhabditis el...		157	7.2e-13	1
gi 3484	(X03579) CDC25 protein (aa 1-1588) {...		136	3.4e-12	3
sp P04821 CC25_YEAST	CELL DIVISION CONTROL PROTEIN 25 /pi...		136	3.4e-12	3
gi 915328	(U24070) Munc13-1 [Rattus norvegicus]		151	5.5e-12	1
pir  A46199	nucleotide-exchange-factor homolog c...		149	5.6e-12	1
pdb 1PTR	Molecule: Protein Kinase C Delta Ty...		136	1.5e-11	1
gi 915330	(U24071) Munc13-2 [Rattus norvegicus]		150	1.6e-11	2
gi 474982	(D21239) 'C3G protein' [Homo sapiens...		131	3.3e-11	3
gi 1763306	(U75361) Munc13-3 [Rattus norvegicus]		153	6.4e-11	2
gi 806957	guanine-nucleotide exchange factor C...		128	7.8e-11	3
sp Q03385 GNDS_MOUSE	GUANINE NUCLEOTIDE DISSOCIATION STIM...		133	1.0e-10	2
pir  BVBYL1	LTE1 protein - yeast (Saccharomyces ...		139	1.9e-10	1
gi 452242	(D21354) a putative guanine nucleoti...		139	2.7e-10	1
sp P07866 LTE1_YEAST	LOW TEMPERATURE ESSENTIAL PROTEIN /p...		139	2.7e-10	1
gi 509050	(Z22521) protein kinase C delta (Hom...		137	4.0e-10	1
gi 520587	(D10495) protein kinase C delta-type...		137	4.6e-10	1
sp P05130 KPC1_DROME	PROTEIN KINASE C, BRAIN ISOZYME (PKC...		137	4.7e-10	1
pir  S35704	protein kinase C (EC 2.7.1.-) delta ...		137	4.7e-10	1
sp Q05655 KPCD_HUMAN	PROTEIN KINASE C, DELTA TYPE (NPKC-D...		137	4.7e-10	1
pir  S40279	protein kinase C mu - human /pir  A5...		137	4.9e-10	1
sp P09215 KPCD_RAT	PROTEIN KINASE C, DELTA TYPE (NPKC-D...		135	9.0e-10	1
gi 520878	(Z34524) serine/threonine protein ki...		133	1.8e-09	1
gi 1519719	(U68142) RalGDS-like [Homo sapiens]		115	3.8e-09	3

12/32

## FIGURE 13(a) (i)

MCG7 - Cloning of a novel human gene that encodes a guanine exchange factor

CGATTTCATTCTCGCTCCCCACAGGTCCTCTCCCCAAAATATTCCTATCTTGTCCTAG 60  
I S F L A P H R S L S P K Y S H L V L 19  
CCCATCCCCAGACTATCTCAAGGACCAGCTGTCCCCACGCCCCGACCTCCACTAGGCC 120  
A H P P D Y L K D Q L S P R P R P P L G 39  
TGTGCCACCCGCTGCCTGCAGGAAGACGCCCCGTCCCGGGCCGGGTAGCCCCATGGGAA 180  
L C H P L P A G R R P V P G R V S P M G 59  
CGcagcgccctgtgtggccgcgggactcaaggctggcctggctcaagtgaacagcacgtcc 240  
T Q R L C G R G T Q G W P G S S E Q H V 79  
aggaggcgacctcgtccgcggggtttgcattctgggggtggacgagctggGGGTTCGGTCCG 300  
Q E E A T S S A G L H S G V D E L G V R S 99  
AGCCCCGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCAGCCCACCCCGCGCCGGCGGCCA 360  
E P G G R L P E R S L G P A H P A P A A 119  
TGGCAGGCACCCTGGACCTGGACAAGGCTGCACGGTGGAGGAGCTGCTCCGCGGGTGCA 420  
M A G T L D L D K G C T V E E L L R G C 139  
TCGAAGCCTTCGATGACTCCGGGAAGGTGCGGGACCCGAGCTGGTGCGCATGTTCTCA 480  
I E A F D D S G K V R D P Q L V R M F L 159  
TGATGCACCCCTGGTACATCCCCTCTCTCAGCTGGCGGCCAAGCTGCTCCACATCTACC 540  
M M H P W Y I P S S Q L A A K L L H I Y 179  
AACAATCCCGGAAGGACAACCTCCAATTCCCTGCAGGTGAAAACGTGCCACCTGGTCAGGT 600  
Q Q S R K D N S N S L Q V K T C H L V R 199  
ACTGATCTCCGCTTCCCGAGCGGAGTTTGACTTGAACCCGGAGTTGGCTGAGCAGATCA 660  
Y W I S A F P A E F D L N P E L A E Q I 219  
AGGAGCTGAAGGCTCTGCTAGACCAAGAAGGGAACCGACGGCACAGCAGCCTAATCGACA 720  
K E L K A L L D Q E G N R R H S S L I D 239  
TAGACAGCGTCCCTACCTACAAGTGAAGCGGCAGGTGACTCAGCGGAACCTGTGGGAC 780  
I D S V P T Y K W K R Q V T Q R N P V G 259  
AGAAAAAGCGCAAGATGTCCCTGTTGTTTGACCACCTGGAGCCCATGGAGCTGGCGGAGC 840  
Q K K R K M S L L F D H L E P M E L A E 279  
ATCTCACCTACTTGGAGTATCGCTCCTTCTGCAAGATCCTGTTTCAGGACTATCACAGTT 900  
H L T Y L E Y R S F C K I L F Q D Y H S 299  
TCGTGACTCATGGCTGCACTGTGGACAACCCCGTCTGGAGCGGTTTCATCTCCCTCTTCA 960  
F V T H G C T V D N P V L E R F I S L F 319  
ACACGCTCTCACAGTGGGTGCAGCTCATGATCCTCAGCAAACCCACAGCCCCGCGAGCGG 1020  
N S V S Q W V Q L M I L S K P T A P Q R 339  
CCCTGGTCATCACACACTTTGTCCACGTGGCGGAGAAGCTGCTACAGCTGCAGAACTTCA 1080  
A L V I T H F V H V A E K L L Q L Q N F 359  
ACACGCTGATGGCAGTGGTCCGGGGCCTGAGCCACAGCTCCATCTCCCGCCTCAAGGAGA 1140  
N T L M A V V G G L S H S S I S R L K E 379  
CCCACAGCCACGTTAGCCCTGAGACCATCAAGCTCTGGGAGGGTCTCACGGAAGTAGTGA 1200  
T H S H V S P E T I K L W E G L T E L V 399  
CGGCGACAGGCAACTATGGCAACTACCGGCGTCCGCTGGCAGCCTGTGTGGGCTTCCGCT 1260  
T A T G N Y G N Y R R R L A A C V G P R 419  
TCCCGATCCTGGGTGTGCACCTCAAGGACCTGGTGGCCCTGCAGCTGGCACTGCCTGACT 1320  
F P I L G V H L K D L V A L Q L A L P D 439  
GGCTGGACCCAGCCCGGACCCGGCTCAACGGGGCAAGATGAAGCAGCTCTTTAGCATCC 1380  
W L D P A R T R L N G A K M K Q L F S I 459  
TGGAGGAGCTGGCCATGGTGACCAGCCTGCGGCCACAGTACAGGCCAACCCCGACCTGC 1440  
L E E L A M V T S L R P P V Q A N P D L 479  
TGAGCCTGCTCACGGTGTCTCTGGATCAGTATCAGACGGAGGATGAGCTGTACCAGCTGT 1500  
L S L L T V S L D Q Y Q T E D E L Y Q L 499  
CCCTGCAGCGGGAGCCGCTCCAAGTCTCGCAACCAGCCCCACGAGTTGCACCCAC 1560  
S L Q R E P R S K S S P T S P T S C T P 519  
CACCCCGGCCCCCGGTACTGGAGGAGTGACCTCGGCTGCCAAACCAAGCTGGATCAGG 1620  
P P R P P V L E E W T S A A K P K L D Q 539  
CCCTCGTGGTGGAGCACATCGAGAAGATGGTGGAGTCTGTGTTCCGGAACCTTTGACGTCG 1680

FIGURE 13(a) (ii)

A L V V E H I E K M V E S V F R N F D V 559  
ATGGGGATGGCCACATCTCACAGGAAGAATTCCAGATCATCCGTGGGAACTTCCCTTACC 1740  
D G D G H I S Q E E F Q I I R G N F P Y 579  
TCAGCGCCTTTGGGGACCTCGACCAGAACCAGGATGGCTGCATCAGCAGGGAGGAGATGG 1800  
L S A F G D L D Q N Q D G C I S R E E M 599  
TTTCCTATTTCTGCGCTCCAGCTCTGTGTTGGGGGGCGCATGGGCTTCGTACACAAC 1860  
V S Y F L R S S S V L G G R M G F V H N 619  
TCCAGGAGAGCAACTCCTTGCGCCCGTCGCTGCCGCCACTGCAAAGCCCTGATCCTGG 1920  
F Q E S N S L R P V A C R H C K A L I L 639  
GCATCTACAAGCAGGGCCTCAAATGCCGAGCCTGTGGAGTGAAGTCCACAAGCAGTGCA 1980  
G I Y K Q G L K C R A C G V N C H K Q C 659  
AGGATCGCCTGTCAAGTTGAGTGTGGGCGCAGGGCCAGAGTGTGAGCCTGGAGGGGTCTG 2040  
K D R L S V E C R R R A Q S V S L E G S 679  
CACCTCACCCTCACCCATGCACAGCCACCATCACCGCGCCTTCAGCTTCTCTCTGCCCC 2100  
A P S P S P M H S H H H R A F S F S L P 699  
GCCCTGGCAGGCGAGGCTCCAGGCCTCCAGAGATCCGTGAGGAGGAGGTACAGACGGTGG 2160  
R P G R R G S R P P E I R E E E V Q T V 719  
AGGATGGGGTGTGTTGACATCCACTTGTAAATAGATGCTGTGGTTGGATCAAGGACTCATTC 2220  
E D G V F D I H L \* 728  
CTGCCTTGGAGAAAATACTTCAACCAGAGCAGGGAGCCTGGGGGTGTGCGGGCAGGAGGC 2280  
TGGGGATGGGGGTGGGATATGAGGGTGGCATGCAGCTGAGGGCAGGGCCAGGGCTGGTGT 2340  
CCCTAAGGTTGTACAGACTCTTGTGAATATTTGTATTTTCCAGATGGAATAAAAAGGCC 2400  
GTGTAATTAACCTTC(A)<sub>n</sub>

14/32

## FIGURE 13(b)

CGATTTCATTCTCGCTCCCCACAGGTCCCTCTCCCCAAAATATTCCCATCTTGTCCTAG 60  
CCCATCCCCCAGACTATCTCAAGGACCAGCTGTCCCCACGCCCCCGACCTCCACTAGGCC 120  
TGTGCCACCCGCTGCCTGCAGGAAGACGCCCCGGTCCCGGGCCGGGTTAGCCCCATGGGAA 180  
CGGGGTTCCGGTCCGAGCCCCGGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCCAGCCCACCC-240  
g v r s e p g g r l p e r s l g p a h p  
CGCGCCGGCGGCCATGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGGAGGAGCT-300  
a p a a M A G T L D L D K G C T V E E L

## FIGURE 14

1 MAGTLDDLKGC...TVEELLRGCI EAF..DDSGKVRDPQLVRMFLMMHPW 45  
 |.:.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:|  
 1 MSSKVEEDQHQELLTEDQLVARCVECFDVDEEDEVEDIEFVDALFLSHQW 50  
 46 YIPSSQLAAKLLHIYQQSRKDNSNSLQVKTCHLVRYWISAFPAEFDLNPE 95  
 .|. | |.:.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:|  
 51 LSDSLSLITHFVNIFYQETRNVQRE...AVCRAVSFWIEKFPMHFDAQPO 97  
 96 LAEIQIKELKALLDQEGNRRHSSLIDIDSVPYKWKQVTRQNPVGQKK.. 143  
 :.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:| |.:.:|  
 98 VCAQVVRKLTIAEDINENIRNGL.DVSALPSFAWLRAVSVRNPLAKQTIV 146  
 144 .....RKMSLLFDHLEPMELAEHLTYLEYR 168  
 :|| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 147 RVDFETLPTPGTPPPFPIASKKFSLTAFSLSFVQASPSDISTSLSHIDYR 196  
 169 SFCKILFQDYHSFVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAP 218  
 :.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 197 VLSRISITELKQYVKDGHRLRSCPLERSISVFNNLSNWVQCMILNKTPK 246  
 219 QRALVITHFVHVAEKLLOLONFNTLMAVVGGLSHSSISRLKETHSHVSPE 268  
 :|| :.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 247 ERAEILVKFVHVAKHLRKINNENTLMSVVGGITHSSVARLAKTYAVLSND 296  
 269 TIKLWEGLTELVTATGNYGNYRRRLAAC.VGFRFPILGVHLKDLVALQLA 317  
 . | |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 297 IKKELTQLTNLLSAQHNFCEYRKALGACNKKFRIPITIGVHLKDLVAINCS 346  
 318 LPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPV.QANPDLLSLLTV 366  
 :.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 347 GANFEKT..KCISSDKLVKLSKLLSNFLVFNQKGNLPEMMDLINTLKV 394  
 367 SLDQYQTEDELYQLSLOREPRSKSSPTSPTSCTPPPRPPVLEEWTSAAKP 416  
 ||| :.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 395 SLDIRYNDDIYELSLRREPCTFMN.....FEPsRGLVFAEWASGTV 437  
 417 KLDQALVVEHIEKMVESVFRNFDVDGEGHISOEFOIIRGNFFYLSAFGD 466  
 |.| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 438 APDNATVSKHISAMVDAVFKHYDHD RDGEISOEFOIIRGNFFYLSAFGD 487  
 467 LDONODGCISREEMVSYFLRSS.SVLGGRMGFVHNFOESENRLRPVACRHC 515  
 :| :|| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 488 IDVMDGOISKDELKTYFMAANKNTKDLRRGFKHNFHETTELTPTCNHC 537  
 516 KALILGIYKOGCLKCRACGVNCHKOCKDRLSVECCRRAQSVSLEGSAPSFS 565  
 . |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 538 NKLLWGILROGFKCKDCGLAVHSCCKSNAVAECCRKSSSNLTRA EWFA 587  
 566 PMSHHHRAFSFSLPRPGRRGSRPPEIREEEVQTVEDGVFDIHL 609  
 |. | :| :.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:| |.:|  
 588 PRGSMRSRIINTC....NNSGSTPDEEIGLVSLACEEVFEDDDL 627

FIGURE 15

```

human    CGATTTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTGTCTCTAG 60
human    CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC 120
human    TGTGCCACCC GCTGCCCTGCA GGAAGACGCC CGGTCCCAGG CCGGGTTAGC CCCATGGGAA 180
human    CGCAGCGCCT GTGTGGCCGC GGGACTCAAG GCTGGCCTGG CTCAAGTGAA CAGCACGTCC 240
mouse    ***tcag** ****ag**** t***** **a*g***t>
human    AGGAGGCGAC CTCGTCCGCG GGTTCGCATT CTGGGGTGGA CGAGCTGGGG GTTCGGTCCG 300
                                     acagg
                                     |
mouse    g*****t**a **catt** ***** **aa**aa* g**ct***** **a**aat**>
human    AGCCCGGTGG GAGGCTCCCG GAGCGCAGCC TGGGCCACG CCACCCCGCG CCGGCGGCCA 360
mouse    ***a*t**** *****tga ***t*t*a*t ***t*t*** ***-tg**a *****a*****>
human    TGGCAGGCAC CCTGGACCTG GACAAGGGCT GCACGGTGGA GGAGCTGCTC CGCGGGTGCA 420
mouse    ***ga**** t***** *****t* ****c***** ***** **t**c**t*>
human    TCGAAGCCTT CGATGACTCC GGAAGGTGC GGGACCCGCA GCTGGTGCGC ATGTTCTTCA 480
mouse    *****t*****t **a***** *a**t**a** **a***** *****t*****>
human    TGATGCACCC CTGGTACATC CCCTCCTCTC AGCTGGCGGC CAAGCTGCTC CACATCTACC 540
mouse    *****a ****t***** *****tt* g**a***** **t*****t*>
human    AACAATCCCG GAAGGACAAC TCCAATTCCC TGCAGGTGAA AACGTGCCAC CTGGTCAGGT 600
mouse    *g***** *****t* *a***a**** *****t*** t*****>
human    ACTGGATCTC CGCCTTCCCA GCGGAGTTTG ACTTGAACCC GGAGTTGGCT GAGCAGATCA 660
mouse    *****a***** **a*****c* ***** a***c***** **a*****>
human    AGGAGCTGAA GGCTCTGCTA GACCAAGAAG GGAACCGACG GCACAGCAGC CTAATCGACA 720
mouse    *****t** ***** *****ca* ***** **c*****>
human    TAGACAGCGT
mouse    *c**g**t**

```

17/32

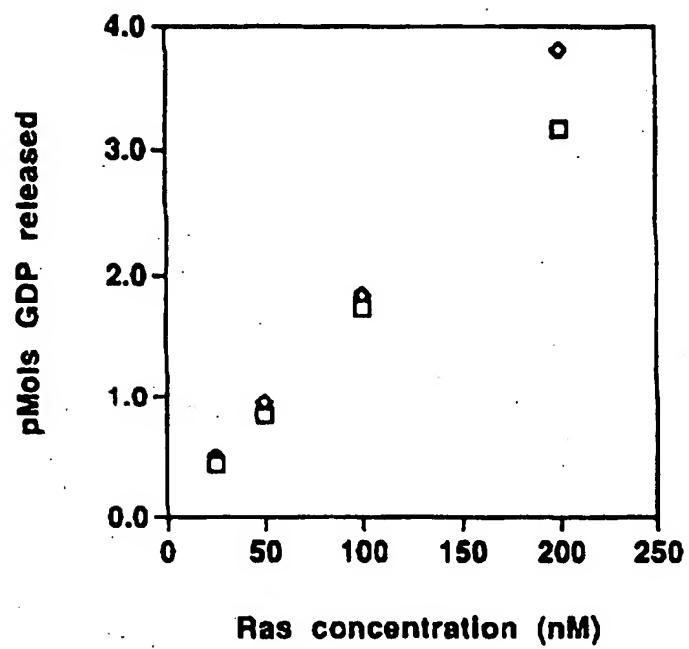
## FIGURE 16

CACGCCTCGGAAGGGAGGTTTGGGGTCGGTGGTTTCACAGTGAGTGTGTCTGAAGCCAAA 60  
TGGTCGGAAACCGTTACCCGCTCTCCTAGGCCCGGCTAGTGGGGACCCCAACCGCCTGCG 120  
\* A R L V G T P T A C>  
GCTGCCCCCTCCCAAGTTCCTCCCTGTTGGCCAGGCATCCAGGTCTCCAGTCTCCGAGCTG 180  
G C P S Q V P P C W P G I Q V S S L R A>  
CGGAGAACCCACCGCCACATGCGGCTGCCCCTTTCCATTGACCCTGTGGGGAGCCAGGC 240  
A E N P P P H A A A P F H S T L W G A R>  
TTCCGGGGCCCCGTTCTCTGTGTGAAGTGGGCCCCCGCCCCATTCCAGACATCAA 300  
L P G P R S S C V N W A P R P H S Q T S>  
GGCCGCGTCTCCAGATAGCCACGATTTTCATTCTCGCTCCCCACAGGTCCCTCTCCCCAA 360  
R P R L Q I A T I S P L A P H R S L S P>  
AATATTCCCATCTTGTCTAGCCCATCCCTCAGACTATCTCAAGGACCAGCTGTCCCCAC 420  
K Y S H L V L A H P P D Y L K D Q L S P>  
GCCCCGACCTCCACTAGGCCTGTGCCACCCGCTGCCTGCAGGAAGACGCCCGGTCCCGG 480  
R P R P P L G L C H P L P A G R R P V P>  
GCCGGGTTAGCCCCATGGGAACGcagcgccctgtgtggccgcgggactcaaggctggcctg 540  
\* p h g n  
G R V S P M G T Q R L C G R G T Q G W P>  
gctcaagtgaacagcacgtccaggagggcgacctcggtccgcgggtttgcattctgggggtgg 600  
G S S E Q H V Q E A T S S A G L H S G V>  
acgagctggGGGTTTCGGTCCGAGCCCGGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCCAG 660  
D E L G V R S E P G G R L P E R S L G P>  
CCCACCCCGCGCCGGCGGCCATGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGG 720  
A H P A P A A M A G T L D L D K G C T V>



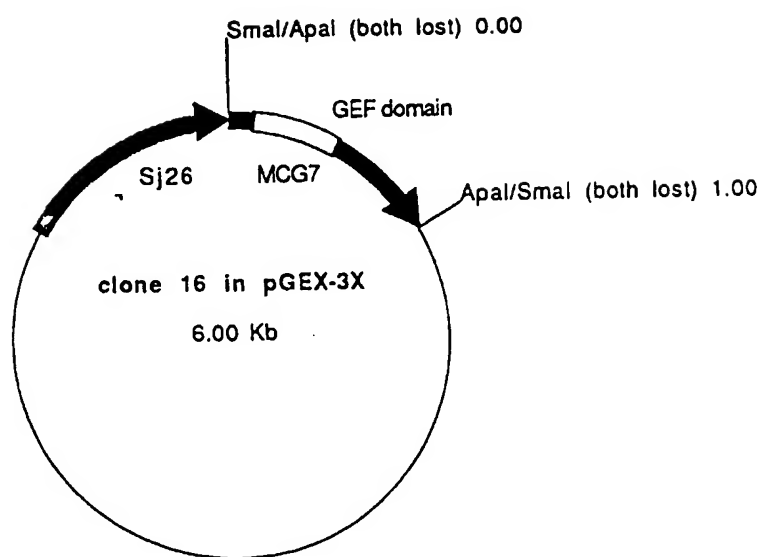
18/32

FIGURE 17



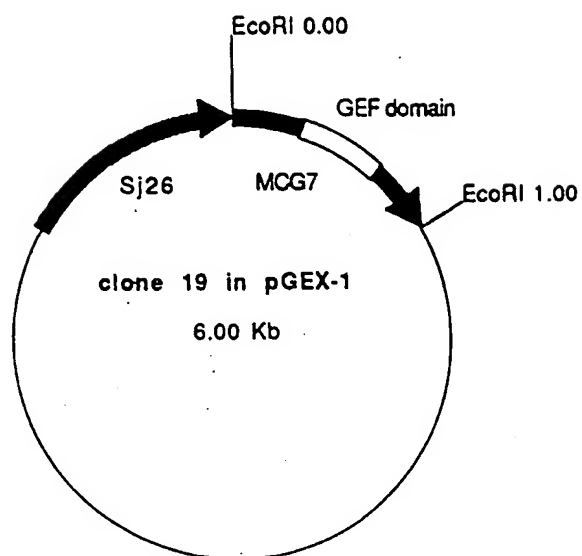
19/32

**FIGURE 18 (Cont. I)**



**Plasmid name:** clone 16 in pGEX-3X

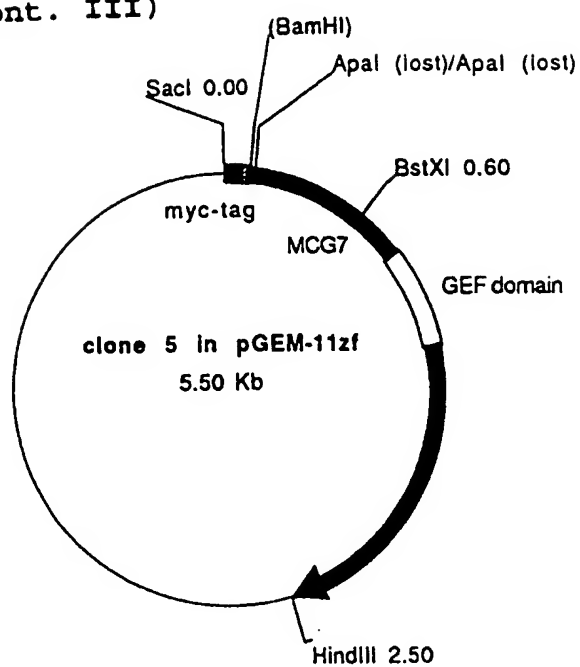
**Plasmid size:** 6.00 kb

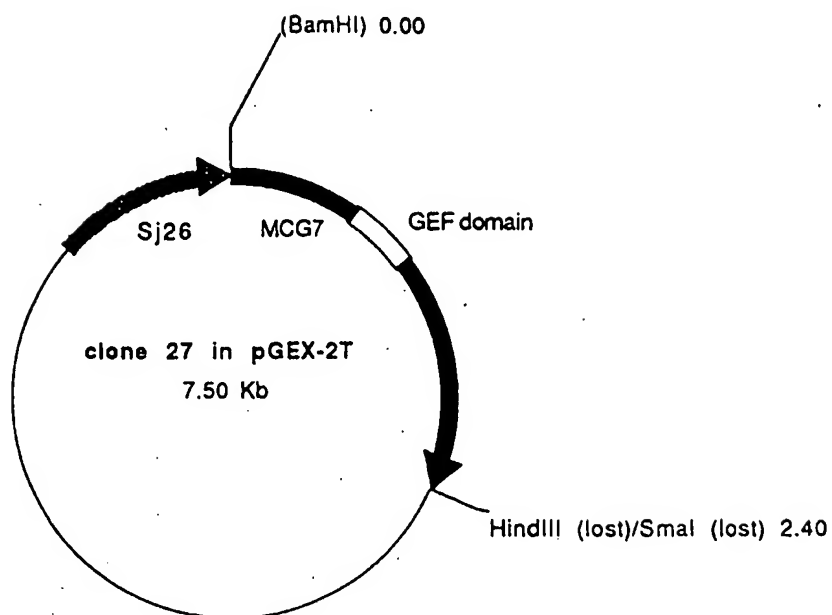
**FIGURE 18 (Cont. II)**

**Plasmid name:** clone 19 in pGEX-1

**Plasmid size:** 6.00 kb

21/32

**FIGURE 18 (Cont. III)****Plasmid name:** clone 5 in pGEM-11zf**Plasmid size:** 5.50 kb

**FIGURE 18 (Cont. IV)**

**Plasmid name:** clone 27 in pGEX-2T

**Plasmid size:** 7.50 kb

23/32

FIGURE 19

GCCCCGCCGCC															ATG	CCG	CCC	TTA	CTG	CCC	CTG	CGC	CTG	TGC	CGG	CTG	TGG	49
															Met	Pro	Pro	Leu	Leu	Pro	Leu	Arg	Leu	Cys	Arg	Leu	Trp	
															1				5					10				
CCC	CGC	AAC	CCT	CCC	TCC	CGG	CTC	CTC	GGA	GCG	GCC	GCC	GGG	CAG	CGG	97												
Pro	Arg	Asn	Pro	Pro	Ser	Arg	Leu	Leu	Gly	Ala	Ala	Ala	Gly	Gln	Arg													
															15			20		25								
TCC	AGA	CCC	AGT	ACT	TAT	TAT	GAA	CTG	TTG	GGG	GTG	CAT	CCT	GGT	GCC	145												
Ser	Arg	Pro	Ser	Thr	Tyr	Tyr	Glu	Leu	Leu	Gly	Val	His	Pro	Gly	Ala													
															30			35		40		45						
AGC	ACT	GAG	GAA	GTT	AAA	CGA	GCT	TTC	TTC	TCC	AAG	TCC	AAA	GAG	CTG	193												
Ser	Thr	Glu	Glu	Val	Lys	Arg	Ala	Phe	Phe	Ser	Lys	Ser	Lys	Glu	Leu													
															50			55		60								
CAC	CCA	GAC	CGG	GAC	CCT	GGG	AAC	CCA	AGC	CTG	CAC	AGC	CGC	TTT	GTG	241												
His	Pro	Asp	Arg	Asp	Pro	Gly	Asn	Pro	Ser	Leu	His	Ser	Arg	Phe	Val													
															65		70		75									
GAG	CTG	AGC	GAG	GCA	TAC	CGT	GTG	CTC	AGC	CGT	GAG	CAG	AGC	CGC	CGC	289												
Glu	Leu	Ser	Glu	Ala	Tyr	Arg	Val	Leu	Ser	Arg	Glu	Gln	Ser	Arg	Arg													
															80		85		90									
AGC	TAT	GAT	GAC	CAG	CTC	CGC	TCA	GGT	AGT	CCC	CCA	AAG	TCT	CCA	CGA	337												
Ser	Tyr	Asp	Asp	Gln	Leu	Arg	Ser	Gly	Ser	Pro	Pro	Lys	Ser	Pro	Arg													
															95		100		105									
ACC	ACA	GTC	CAT	GAC	AAG	TCT	GCC	CAC	CAA	ACA	CAC	AGC	TCC	TGG	ACA	385												
Thr	Thr	Val	His	Asp	Lys	Ser	Ala	His	Gln	Thr	His	Ser	Ser	Trp	Thr													
															110		115		120		125							
CCC	CCC	AAC	GCA	CAG	TAC	TGG	TCC	CAG	TTT	CAC	AGC	GTG	AGG	CCA	CAG	433												
Pro	Pro	Asn	Ala	Gln	Tyr	Trp	Ser	Gln	Phe	His	Ser	Val	Arg	Pro	Gln													
															130		135		140									
GGG	CCC	CAG	TTG	AGG	CAG	CAG	CAA	CAC	AAA	CAA	AAC	AAA	CAA	GTG	CTG	481												
Gly	Pro	Gln	Leu	Arg	Gln	Gln	Gln	His	Lys	Gln	Asn	Lys	Gln	Val	Leu													
															145		150		155									
GGG	TAC	TGC	CTC	CTC	CTC	ATG	CTG	GCG	GGC	ATG	GGC	CTG	CAC	TAC	ATT	529												
Gly	Tyr	Cys	Leu	Leu	Leu	Met	Leu	Ala	Gly	Met	Gly	Leu	His	Tyr	Ile													
															160		165		170									
GCC	TTC	AGG	AAG	GTG	AAG	CAG	ATG	CAC	CTT	AAC	TTC	ATG	GAT	GAA	AAG	577												
Ala	Phe	Arg	Lys	Val	Lys	Gln	Met	His	Leu	Asn	Phe	Met	Asp	Glu	Lys													
															175		180		185									
GAT	CGG	ATC	ATC	ACA	GCC	TTC	TAC	AAC	GAA	GCC	CGG	GCA	CGG	GCC	AGG	625												

**FIGURE 19 (continued)**

Asp Arg Ile Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg	
190 195 200 205	
GCC AAC AGA GGC ATC CTT CAG CAG GAG CGA CAA CGG CTA GGG CAG CGG	673
Ala Asn Arg Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg	
210 215 220	
CAG CCG CCA CCA TCC GAG CCA ACC CAA GGC CCC GAG ATC GTG CCC CGG	721
Gln Pro Pro Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg	
225 230 235	
GGC GCC GGC CCC TGA GGGGCTC ACCTGGATGG GGCCTGCAGT GCGTTCCCGC	773
Gly Ala Gly Pro *	
240	
TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTGCGAG	832

**FIGURE 20**

```
>sp|P08622|DNAJ_ECOLI DNAJ PROTEIN >pir||HHECDJ heat shock protein dnaJ -  
Escherichia coli >gi|145769 (M12565)-heat shock protein dnaJ  
[Escherichia coli] >gi|216441 (D19483) dnaJ protein [Escherichia  
coli]  
Length = 376
```

```
Score = 138 (63.7 bits), Expect = 1.2e-10, P = 1.2e-10  
Identities = 25/62 (40%), Positives = 39/62 (62%)
```

```
Query: 35 YYELGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLREQSRRS 94  
YYE+LGV A E+++A+ + + HPDR+ G+ ++F E+ EAY VL+ Q R +  
Sbjct: 6 YYEILGVSKTAEEREIRKAYKRLAMKYHPDRNQGDKAEAKFKIKEAYEVLTDQKRAA 65  
  
Query: 95 YD 96  
YD  
Sbjct: 66 YD 67
```



## FIGURE 21

>gi|1703590 (U80439) contains similarity to a DNAJ-like domain (Caenorhabditis elegans)  
Length = 345

Score = 98 (45.2 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12  
Identities = 17/37 (45%), Positives = 28/37 (75%)

Query: 28 QRSRPSTYYELGVPAGASTEELVKRAFFSKSKELHPD 64  
++ R T+YE+LGV A+ E+K AF+++SK++HPD  
Sbjct: 22 KKIRQRTHYEVLGVSTATLSEIKSAFYAQSKVHPD 58

Score = 74 (34.1 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12  
Identities = 17/32 (53%), Positives = 19/32 (59%)

Query: 71 SLHSRFVELSEAYRVLSREQSRRSYDDQLRSG 102  
S + F+EL AY VL R RR YD QLR G  
Sbjct: 64 SATASFLELKNAYDVLRRPADRRLYDYQLRGG 95

Score = 39 (18.0 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12  
Identities = 10/42 (23%), Positives = 19/42 (45%)

Query: 162 LLMLAGMGLHYIAFRKVKQMHLNFMDEKDRITAFYNEARAR 203  
L+++AG Y+ Q L+ + ++D I F + R  
Sbjct: 158 LVLVAGYNGGYLYLLAYNQQLDKLIDEDEIAKCFLRQKEFR 199

>gnl|PID|e281266 (Z81030) C01G10.12 (Caenorhabditis elegans)  
Length = 191

Score = 96 (44.3 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09  
Identities = 17/41 (41%), Positives = 27/41 (65%)

Query: 35 YYELGVPAGASTEELVKRAFFSKSKELHPDRDPCNPSLHSR 75  
YYE++GV A+ +E++ AF K+K+LHPD+ + SR  
Sbjct: 19 YYEIIIGVSASATRQEIRDAFLKKTQQLHPDQSRKSSKSDSR 59

Score = 54 (24.9 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09  
Identities = 10/22 (45%), Positives = 15/22 (68%)

Query: 75 RFVELSEAYRVLSREQSPPSYD 96  
+F+ + EAY VL E+ R+ YD  
Sbjct: 71 QFMLVKEAYDVLNMEERKEVD 92

Score = 35 (16.1 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09  
Identities = 9/44 (20%), Positives = 22/44 (50%)

Query: 141 QGPQLRQQHKKQKQVLGYCLLLMLAGMGLHYIAFRKVKQMHLN 184  
+ P+ + KQ ++L ++A +G + + RK++ L+  
Sbjct: 145 RNPEDEYLREKQKNRMLVLAATVMALICANTVYIRKLQADRLS 188

**FIGURE 22**

>sp|Q10209|YAY1\_SCHPO HYPOTHETICAL 44.8 KD PROTEIN C4H3.01 IN CHROMOSOME I  
 >gi|1184014 (Z69380) unknown [Schizosaccharomyces pombe]  
 Length = 392

Score = 84 (38.8 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08  
 Identities = 13/35 (36%), Positives = 25/36 (69%)

Query: 35 YYELGVHPGASTEIVKRAFFSKSKELHPDRDPGNP 70  
 YY+LLG+ A+ ++K+A+ + + HPD++P +P  
 Sbjct: 9 YYDLLGISTDATAVDIKKAYRKLAVKYHPDKNPDDP 44

Score = 64 (29.5 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08  
 Identities = 14/40 (35%), Positives = 23/40 (57%)

Query: 75 RFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPPTTVHD 114  
 +F ++SEAY+VL E+ R YD + + P+ T +D  
 Sbjct: 50 KFQKISEAYQVLGDEKLRSQYDQFGKEKAVPEQGFTDAYD 89

Score = 37 (17.1 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08  
 Identities = 9/29 (31%), Positives = 15/29 (51%)

Query: 190 DRIITAFYNEARARARANGILQQRQRL 218  
 DR A E A A+ + +++ RQR+  
 Sbjct: 149 DRKQNAQIREREALAKREQEMIEDRRQRI 177

Score = 33 (15.2 bits), Expect = 0.00081, Sum P(3) = 0.00081  
 Identities = 8/19 (42%), Positives = 11/19 (57%)

Query: 140 PQGPQLRQQQHKCNQVLG 158  
 PQG + Q+ + QVLG  
 Sbjct: 44 PQGASEKFKKISEAYQVLG 62

**FIGURE 23**

>gnl|PID|e253406 (X77635) tumorous imaginal discs [Drosophila virilis]  
 >gnl|PID|e263866 (Y07700) Tid58 protein [Drosophila virilis]  
 Length = 529

Score = 153 (70.6 bits), Expect = 9.7e-13, P = 9.7e-13  
 Identities = 27/71 (38%), Positives = 44/71 (61%)

Query: 26 AGQSRSPSTYYELGVHPGASTEIVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRV 85  
 + R + YY LGV A+ +++K+A++ +K+ HPD + +P +F ++SEAY V  
 Sbjct: 72 SSSRMQAKDYATLGVAIGNANAKDIKKAYYELAKKYHPDTNKDDPDASKKQDVSEAYEV 131

Query: 86 LSREQSRRSYD 96  
 LS +Q RR YD  
 Sbjct: 132 LSDDQKRREYD 142

## FIGURE 24

28/32

```

MCG18 -----MPPLLPLRLCRLWP-RN--PP-----SRLLGAA
HDJ-2  MVKETTYVDVLGVKPNATQEELKKAYRKLALKYHPDKN--PN---EGEGFKQISQAYEV
HDJ-1  MGKD--YYQTLGLARGASDEEIKRAYRRQALRYHPDKNKEPG---AEEKFKEIAEAYDV
HSJ1    M-AS--YYEILDVPRASADDIICKAYRRKALQWHPDKN--PDNKEFAEKKFKEVAEAYEV
      . . .

MCG18  AGQSRSPSTY--YELLG VH-----PGA-----ST-EEVKRAFFS--
HDJ-2  LSDAKKRELYDKGGEQA IK-----EGGAGGG-----FGSPMDIFDMFFGGG
HDJ-1  LSDPRKREIFDRYGEGLKGS GP-----SGGSGGGANGTSFSYTFHGDPHAMFAEFFG--
HSJ1    LSDPKHREIYDRYGREGLTGTGTGPSRAEAGSGGP--G--FTFT-FKSPPEVFREFFG--
      . . .

MCG18  KSKELHPORDPGNP---SLHSRFVELSEAYRVLSREQSRRS--YDQQLRSGSPKSPRT
HDJ-2  GRMQRERRGKNNVHQLSVTLEDLYNGATRKALQKNVICDKCEGRGGKKGAVECCPNCRG
HDJ-1  GRNPFDTFFGQRNGEEGMDIDDPFSGFPMGMOGFTNVNFGRS--RSAQEPARKKQDPVPT
HSJ1    SGDPFAELFDDLGP--FSELQNRGSRHSGPFTTSSSFPGHSDFSSSSFSPGAGAFRS
      . . .

MCG18  TVHDKSAHQTHSSWTPPNAQY---WSQFHSVRPQ-----GP-----QLRQQQHKQN
HDJ-2  TGMQIRIHIQIGPMVQQIQSVCMQCQHGERISPK-DRCKSCNGRKIVREKKILEVHIDK
HDJ-1  HDLRVSLEEIYSGCTKGMK-----ISH-KRLNP--D-----GKSIRNEDKILTIEVKK
HSJ1    VSTSTTFVQGRITTRIME---NGQ-ERVEVEED-----GQ-----LKSVTINGVPD
      . . .

MCG18  KQVLGYCLLL-----MLAGMGLHYIAFRKVKQMHLMFDE-KDRIITAFYNearARARAN
HDJ-2  GMKQGQKITTFHGEDQEPGLEPGDIIIVLDQKDHAVFTRRGEDLFMCMDIQLVEALCGFQ
HDJ-1  GWKEGTKITFPKEGDQTSNNIPADIVFVLKDKPHNIFKRDGSDVIYPARISLREALCGCT
HSJ1    DLARGLELSR-RE--QQP-SVTSRSGGTQVQQTASCPLD-SLSEDEDLQLAMAYSLSE
      . . .

MCG18  RGILQQERQRLGQRQPP-PSEPTQGPETVPRGAP-----
HDJ-2  KPISTLDNRTIVITSHPGQIVKHGDIKCVLNEGMPYRRPYEKGRLLIEFKVNFENGFL
HDJ-1  VNVPTLDGRTIPVVF--DVIRPGMRKVPGEGLPLPKTPEKRGDLIEFEVIFPER--I
HSJ1    MEAAGKKPAGGREAQHR-RQGRPRPSTKIQA WGGP--RR--VRG--VKQPNVHPQR-RR
      . . .

MCG18  -----
HDJ-2  SPDKLSLLEKLLPERKEVEETDEM DQVELVDFDPNQERRRHYNGEAYEDDEHFRGGVQC
HDJ-1  PQTSTRVLEQVLPI-----
HSJ1    PLAASSSEHRAQPD-----LIQILTGGSDSLWEEKRGVS-----

MCG18  ---
HDJ-2  QTS
HDJ-1  ---
HSJ1    ---

```

\* = amino acid identity in all 4 proteins

. = conservative substitution

## FIGURE 25

CAAGGAGCCTCTGCCTGCCCCGTCGTCGTCATGCCGTCCTGTTGCTCCAGCTGCCCCCTGC 60  
M P S L L L Q L P L 10  
GCCTATGCCGGCTGTGGCCGCATAGCCTTTCCATCCGACTTCTCACAGCCGCCACAGGGC 120  
R L C R L W P H S L S I R L L T A A T G 30  
AGCGGTCTGTCCCTACTAATTACTATGAATTGTTGGGCGTCATCCGGGTGCCAGCGCTG 180  
Q R S V P T N Y Y E L L G V H P G A S A 50  
AAGAGATTAAACGTGCTTTTTTCACCAAGTCAAAAGAGCTACACCCTGATCGAGACCCTG 240  
E E I K R A F F T K S K E L H P D R D P 70  
GGAACCCAGCCCTGCATAGCCGCTTTGTGGAGCTGAATGAGGCATATCGAGTGCTCAGTC 300  
G N P A L H S R F V E L N E A Y R V L S 90  
GTGAGGAAAGTCGTCGTAACTATGACCACCAGCTGCATTGAGCCAGTCCTCCAAAGTCTT 360  
R E E S R R N Y D H Q L H S A S P P K S 110  
CAGGGAGCACAGCCGAGCCTAAGTATACGCAACAGACACACAGCAGCTCCTGGGAACCCC 420  
S G S T A E P K Y T Q Q T H S S S W E P 130  
CCAACGCTCAATACTGGGCCCAGTTCCACAGTGTGAGGCCGAGGGCCGGAGTCAAGGA 480  
P N A Q Y W A Q F H S V R P Q G P E S R 150  
AGCAGCAGCGTAAACACAACCAGCGGGTCTGCGCTCCTGCTCATGGTGGCAG 540  
K Q Q R K H N Q R V L G Y C L L L M V A 170  
GCATGGGCCTGCACTATGTTGCCTTCAGGAAGCTGGAGCAGGTGCATCGCAGCTTCATGG 600  
G M G L H Y V A F R K L E Q V H R S F M 190  
ATGAAAAGGACCGGATCATTACAGCCATCTACAATGACACTCGGGCCAGGGCCAGGGCCA 660  
D E K D R I I T A I Y N D T R A R A R A 210  
ACAGAGCCAGGATTCAGCACGAGCGCCACGAGAGGCAGCAGCCTCGGGCAGAACCCCTCCC 720  
N R A R I Q Q E R H E R Q Q P R A E P S 230  
TGCCTCCAGAAAGCTCCAGGATCATGCCCCAGGACACAAGCCCCCTGAGAGGCTTAACTAA 780  
L P P E S S R I M P Q D T S P \* 245  
ATGGGACCTTCATTGGTCCTCTCCCTGCTGCCTGTCCAGAACTACACGTGCAATAAACTC 840  
ATTTTCAG(A)n 849

30/32

## FIGURE 26

human MCG18	MPPLL---PLRLCRLWPRNPPSRLLGAAAGQSRPSTYYELLGVHPGASTE EVKRAFFSK
mouse MCG18	MPSLLLQLPLRLCRLWPHSL SIRLLTAATGQRSVPTNYELLGVHPGASAEIKRAFFTK
	.. .. * .....
human MCG18	SKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRTTVHDKSA
mouse MCG18	SKELHPDRDPGNPALHSRFVELNEAYRVLSREESRRNYDHLHSASPPKSSGSTAEFKYT
	..... * .....
human MCG18	.HQTHSS-WTPPNAQYWSQFHSVRPQGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFR
mouse MCG18	QQTHSSSWEPNNAQYWAQFHSVRPQGPESRKQQRKQNRVLYCLLLMVAGMGLHYVAFR
	..... * .....
human MCG18	KVKQMHLNFMDEKDRIITAFYNEARARARANRGILQOERQRLGQRQPPSEPTQGPE---
mouse MCG18	KLEQVHRSFMDEKDRIITATYNDTRARARANRARIQOER---HERQQPRAEPSLPPESSR
	.. * .....
human MCG18	IVPRGAGP
mouse MCG18	IMPQDTSP
	..*..*

31/32

**FIGURE 27**

```
ttgaagtctagccccatcctgggtccaatgcgctcttggtagcctccttccccagctgccc 60
* S L A P S W S N A L L V A S F P S C P

gcccgcggccATGCCGCCCTTACTGCCCCCTGCGCCTGTGCCGGCTGTGGCCCCGCAACCC 120
P A A M P P L L P L R L C R L W P R N P>
```

32/32

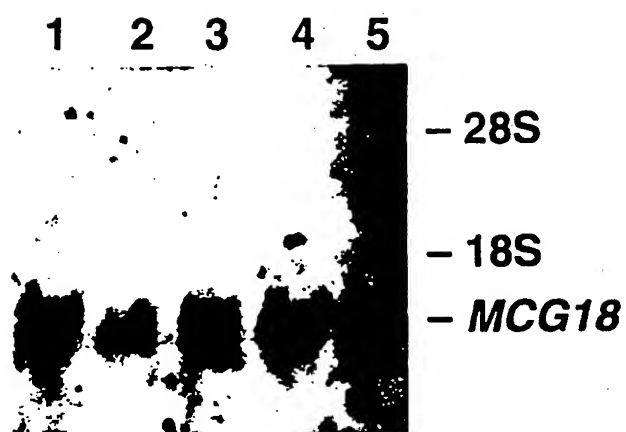
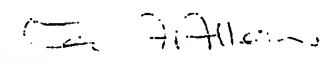


FIGURE 28

# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/AU 98/00380

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>																						
Int Cl <sup>6</sup> : C12N 15/12; C07K 14/47; C07K 16/18; G01N 33/53																						
According to International Patent Classification (IPC) or to both national classification and IPC																						
<b>B. FIELDS SEARCHED</b>																						
Minimum documentation searched (classification system followed by classification symbols) I/C: WPAT (D gene) Sequences provided by Applicant																						
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched																						
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) : EMBL, Genbank, Swiss Prot and PIR: Sequences provided by applicant																						
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>																						
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.																				
P,X	Kedra D, Seroussi E, Fransson I, Trifunovic J, Clark M, Lagercranz J, Blennow E, Mehlin H, Dumanski J, Human Genetics, October 1997 100(5-6) 611-619 The germinal centre kinase gene and a novel CDC25-like gene are located in the vicinity of the PYGM gene on 11q13 EMBL AC Y12339	1-3,8-10,15-18																				
P,X	Guru S C, Agarwal S K, Manickain P, Olufemi S E, et al Genome Research, July 1997 7(7) 725-735. A transcript map for the 2.8-Mb region containing the multiple endocrine neoplasia type I locus TREMBL AC 014616	1, 4-5, 8, 11-12, 15, 19-21																				
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C <input type="checkbox"/> See patent family annex																						
<p>* Special categories of cited documents:</p> <table border="0"> <tr> <td>"A"</td> <td>document defining the general state of the art which is not considered to be of particular relevance</td> <td>"T"</td> <td>later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>"E"</td> <td>earlier document but published on or after the international filing date</td> <td>"X"</td> <td>document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>"L"</td> <td>document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>"Y"</td> <td>document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>"O"</td> <td>document referring to an oral disclosure, use, exhibition or other means</td> <td>"&amp;"</td> <td>document member of the same patent family</td> </tr> <tr> <td>"P"</td> <td>document published prior to the international filing date but later than the priority date claimed</td> <td></td> <td></td> </tr> </table>			"A"	document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	"E"	earlier document but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family	"P"	document published prior to the international filing date but later than the priority date claimed		
"A"	document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention																			
"E"	earlier document but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone																			
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art																			
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family																			
"P"	document published prior to the international filing date but later than the priority date claimed																					
Date of the actual completion of the international search 16 July 1998		Date of mailing of the international search report 20 JUL 1998																				
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929		Authorized officer  GILLIAN ALLEN  Telephone No.: (02) 6283 2266																				



# INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 98/00380

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 1, 2, 4, 6  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
They are to known groups of proteins and lack distinguishing features which would enable a meaningful search.
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Invention 1, defined by claims 2, 3, 9, 10, 16-18, is to nucleotide sequences, amino acid sequences and proteins with a zinc finger domain.

Invention 2, defined by claims 4, 5, 11, 12, 19-21, is to nucleotide sequences and amino acid sequences and proteins which are guanine exchange factors.

Invention 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins which are heat shock proteins or heat shock binding proteins.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 98/00380

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EMBL AC AF012106 DT 6 November 1997 Lloyd S E and Thakker R V DE Homo Sapiens DnaJ protein (HSPF <sub>2</sub> )mRNA, complete cds	1,6-8,13- 15,22-24
P,X	EMBL AC AF 036875 DT 20 May 1998 Silins G, Grimmond S, Hayward N DE Mus musculus multiple endocrine neoplasia type I candidate protein number 18 mRNA, complete cds	1,6-8,13- 15,22-24



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/12, C07K 14/47, 16/18, G01N 33/53</b>		<b>A1</b>	(11) International Publication Number: <b>WO 98/53061</b>
			(43) International Publication Date: 26 November 1998 (26.11.98)
(21) International Application Number: <b>PCT/AU98/00380</b>		(74) Agents: HUGHES, E., John, L. et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU).	
(22) International Filing Date: 22 May 1998 (22.05.98)			
(30) Priority Data:		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
PO 6972 23 May 1997 (23.05.97) AU PO 6973 23 May 1997 (23.05.97) AU PO 6974 23 May 1997 (23.05.97) AU PP 1458 22 January 1998 (22.01.98) AU PP 1459 22 January 1998 (22.01.98) AU PP 1460 22 January 1998 (22.01.98) AU			
(71) Applicant (for all designated States except US): THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH [AU/AU]; 300 Herston Road, Brisbane, QLD 4029 (AU).		Published With international search report.	
(72) Inventors; and			
(75) Inventors/Applicants (for US only): HAYWARD, Nicholas [AU/AU]; 13 Prince Street, Paddington, QLD 4064 (AU). SILINS, Ginters [AU/AU]; 35 Eppalong Street, The Gap, QLD 4061 (AU). GRIMMOND, Sean [AU/GB]; Medical Research Council, Harwell, Didcot, Oxfordshire OX11 0RD (GB). GARTSIDE, Michael [AU/AU]; 19 Thomas Street, Camp Hill, QLD 4152 (AU). HANCOCK, John [AU/AU]; 141 Airley Road, Pullenvale, QLD 4069 (AU).			
(54) Title: THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN			
(57) Abstract			
<p>The present invention relates generally to three novel human genes with gene regulatory function. These genes encode a zinc finger protein, a guanine nucleotide exchange protein and a heat shock protein or heat shock binding protein. The invention includes derivatives and mammalian animal, insect, nematodes, avian and microbial homologues of these genes. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

## FIELD OF THE INVENTION

5 The present invention relates generally to a novel human gene and its derivatives and to mammalian, animal, insect, nematodes, avian and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

10

## BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

15

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis and in conventional pharmaceutical preparations as well as in gene and protein replacement therapies.

20

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. Molecules of particular interest targeted by the inventors were gene regulators including regulatory proteins, signal transducers and heat shock proteins.

25

Gene expression generally requires interaction between a regulatory protein and an appropriate recognition sequence of a target gene. Regulatory proteins comprise in many cases a domain or motif which facilitates binding to DNA. One particular motif comprises small sequence units repeated in tandem with each unit folded about a zinc atom to form separate structural domains.

30 This motif is now referred to as a zinc finger domain. Such a domain is generally defined by the number of cysteine (C) and histidine (H) residues.

- 2 -

In addition, knowledge of cellular interaction in the control of cell proliferation is essential in the rational design of specific therapeutic strategies aimed at controlling proliferative disorders. Such proliferative disorders including a range of cancers, inflammatory conditions and atherosclerosis. An important aspect of cellular interaction is in signal transduction *via* receptors to intracellular transducers. One key signal transducer is Ras which couples the receptors for diverse extracellular signals to different effectors. Ras directly activates the downstream kinase Raf which in turn induces the mitogen activated protein kinase (MAPK) cascade.

Another regulatory mechanism involves heat shock proteins. The *Escherichia coli* heat shock protein, DnaJ, is the founding member of a family of proteins which are associated with protein folding, protein complex assembly and transit through subcellular components.

Prokaryotic and eukaryotic DnaJ homologues have a modular organisation consisting of a J domain, a glycine-rich spacer, CXXCXGXG [SEQ ID NO:1] repeats and a C-terminal region with no obvious sequence features, as well as additional sequences for protein targeting. The J domain is anticipated to mediate interaction with heat shock 70 proteins (Hsp70) and consists of some 70 amino acids, frequently located at the N-terminus of the protein.

In accordance with the present invention, a genes have been identified from the human genome which encodes proteins having a regulatory role. One gene, in accordance with the present invention encodes a protein with an N-terminal region resembling a zinc-finger domain of a novel type. Another gene encodes a protein involved in guanine nucleotide exchange factor (GEF) signalling pathways. Yet another gene encodes a protein which is a heat shock protein or heat shock-like protein which may have a role in tumour suppression.

25

#### SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Sequence identity numbers (SEQ ID NOs.) for nucleotide and amino acid sequences referred to in the subject specification are defined after the bibliography. A summary of SEQ ID NOs. is also given in Table 1.

- 5 One aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
- 10 Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an  $(\text{HC}_3)_2$  type.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule  
15 comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence  
20 of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:2 defines the gene, *mcg4*. This gene encodes  
25 a product, MCG4, having an amino acid sequence set forth in SEQ ID NO:3.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg4* gene portion, which *mcg4* gene portion is capable of encoding an MCG4 polypeptide or a  
30 functional or immunologically interactive derivative thereof.

- 4 -

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion  
5 and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or  
10 multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological  
15 sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

A further aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an  
20 amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

25

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence  
30 of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the



- 5 -

nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:4 or 6 defines the gene, *mcg7*. This gene encodes a product, MCG7, having an amino acid sequence set forth in SEQ ID NO:5 or 7.

5

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an MCG7 polypeptide or a functional or immunologically interactive derivative thereof.

10

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion  
15 and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or  
20 multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological  
25 sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an  
30 amino acid sequence having homology to a heat shock protein or a heat shock binding protein or a derivative thereof.

Another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 5 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 41°C to the nucleotide sequence set forth in (i), (ii) or (iii).

10

The nucleotide sequence set forth in SEQ ID NO:8 defines the gene, *mcg18*. This gene encodes a product, MCG18, having an amino acid sequence set forth in SEQ ID NO:7.

Even yet another aspect of the present invention provides a genetic construct comprising a vector  
15 portion and an animal, more particularly a mammalian and even more particularly a human *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition  
20 caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

25

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

30

Another aspect of the present invention contemplates a method for detecting MCG18 or a

- 7 -

derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

5

A summary of SEQ ID Nos. referred to in the subject specification is shown in Table 1.

**TABLE 1**  
**SUMMARY OF SEQ ID Nos.**

5	SEQ ID NO.	DESCRIPTION
	1	amino acid repeat sequence in DnaJ homologues
	2	Nucleotide sequence of <i>mcg4</i>
	3	amino acid sequence of MCG4
	4	nucleotide sequence of <i>mcg7</i>
10	5	amino acid sequence of MCG7
	6	nucleotide sequence of <i>mcg7</i> within exon of nucleotides 183-288
	7	amino acid sequence of MCG7 within exon of nucleotide 183-288
	8	nucleotide sequence of <i>mcg18</i>
	9	amino acid sequence of MCG18
15	10-18	amino acid sequence identified using BESTFIT
	19	sequence of pGEX and <i>mcg7</i> junction
	20	sequence of pGEX and <i>mcg7</i> junction
	21	nucleotide sequence of <i>myc</i> -tag/ <i>mcg7</i> junction
	22	amino acid sequence corresponding to SEQ ID NO:21
20	23	nucleotide sequence of pGEX and <i>mcg7</i> junction
	24	amino acid sequence corresponding to SEQ ID NO:23
	25-36	<i>mcg7</i> -specific oligonucleotide
	37-45	<i>mcg18</i> -specific oligonucleotide

25 Single and three letter abbreviations for amino acid residues are shown in Table 2.

- 9 -

TABLE 2

Amino Acid	Three-letter Abbreviation	One-letter Symbol
5		
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
10 Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
15 Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
20 Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
25 Valine	Val	V
Any residue	Xaa	X

**BRIEF DESCRIPTION OF THE FIGURES**

**Figure 1** is a representation of the nucleotide sequence [SEQ ID NO:2] and corresponding amino acid sequence [SEQ ID NO:3] of *mcg4*.

5

**Figure 2** is a representation of the alignment of the human MCG4 amino acid sequence with a translation of a partial murine expressed sequence tag (EST).

**Figure 3** is a representation of the alignment of the human MCG4 amino acid sequence with a  
10 translation of a partial nematode EST.

**Figure 4** is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn represent zinc atoms.

15

**Figure 5** is a representation of sensitive sequence homology search of related cysteine-containing motifs in another *Caenorhabditis elegans* protein.

**Figure 6** is a representation showing that a related cysteine containing motif is present in the  
20 GATA-binding transcription factor from *Saccharomyces pombe*.

**Figure 7** is a Northern blot showing expression of *mcg4* in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15 $\mu$ g total RNA derived from various human cancer cell lines. Lanes 1-5, respectively, contain RNA from H69 lung  
25 carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

**Figure 8** is a representation of a partial alignment of *mcg4* with human ESTs AA074703 and AA134788.

30

**Figure 9** is a representation of the partial nucleotide sequence alignment between a human

- 11 -

(W32939) and mouse (AA242159) *mcg4*-like EST in the putative 5' UTR of the *mcg4* cDNA. The putative initiation codon is underlined and the region upstream represents 5' UTR.

**Figure 10** is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

**Figure 11** is a diagrammatic representation of the domains of MCG4

zinc finger consensus: CX<sub>2</sub>HX<sub>4</sub>CX<sub>2</sub>CX<sub>4</sub>HX<sub>2</sub>CX<sub>17</sub>CX<sub>2</sub>CX<sub>18</sub>HX<sub>2</sub>CX<sub>18</sub>CX<sub>2</sub>C

acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged

10 basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged

leucine zipper domain consensus: LX<sub>6</sub>LX<sub>6</sub>RX<sub>6</sub>LX<sub>6</sub>L

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa261) LX<sub>6</sub>LXLX<sub>6</sub>LXLX<sub>6</sub>L (aa 286).

15 **Figure 12** is a representation showing similarity of MCG7 with GEFs of various organisms.

**Figure 13(a)** is a representation of the nucleotide sequence [SEQ ID NO:4] and corresponding amino acid sequence [SEQ ID NO:5] of *mcg7*. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

20

**Figure 13(b)** is a representation of the partial nucleotide sequence [SEQ ID NO:6] and corresponding amino acid sequence [SEQ ID NO:7] of *mcg7* but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon  
25 of nucleotides 183-288.

**Figure 14** is a representation showing a comparison between MCG7 and a homologue from *Caenorhabditis elegans* using the BESTFIT algorithm. In the figure, the following sequences are underlined:

30

EF-Hand= PROSITE DATABASE NO. PD0C00018

- 12 -

1a nematode DVDEEDEVEDIEF [SEQ ID NO:10]  
 1b human DVDGDGHISQEEF [SEQ ID NO:11]  
     nematode DHDRDGFISQEEF [SEQ ID NO:12]  
 1c human DQNQDGCISREEM [SEQ ID NO:13]  
 5     nematode DVDMDGQISKDEL [SEQ ID NO:14]

**GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B**

2 human HFVHVAEKLLQLQNFTLMAVVGGLSHSSISRLKETH [SEQ ID NO:15]  
     nematode KFVHVAKHLRKINNFTLMSVVG GITHSSVARLAKTY  
 10                   [SEQ ID NO:16]

**DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379**

3 human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC  
                   [SEQ ID NO:17]  
 15 nematode HNFHETTFLTP TTCNHCKNLLWGILRQGFKCKDCGLAVHSCCKSNAVAEC  
                   [SEQ ID NO:18]

**Figure 15** is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine *mcg7* cDNA (GenBank Acc. No. W71787 and AA237373). The putative  
 20 initiation codon is underlined. The murine sequence represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

25 **Figure 16** is a representation of further 5' nucleotide and corresponding amino acid sequence for human *mcg7*. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Fig. 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.

30 **Figure 17** is a graphical representation of a GDP release assay. □ Experiment #1 (mean of duplicates). ◇ Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols



of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 mins.

Estimated reaction constants:

$K_m = 2.1\mu\text{M}$ ,  $V_{\max} = 37\text{pMol}/6\text{min}/36\text{pMol}$  [Expt#1]

5  $K_m = 1.5\mu\text{M}$ ,  $V_{\max} = 30.3\text{pMol}/6\text{ min}/36\text{pMol}$  [Expt#2]

Figure 18 depicts various recombinant plasmids containing partial or full-length *mcg7*.

Figure 19 is a representation of the nucleotide sequence [SEQ ID NO:8] and corresponding  
10 amino acid sequence [SEQ ID NO:9] of *mcg18*.

Figure 20 is a representation showing that MCG18 has partial homology to *E. coli* DnaJ.

Figure 21 is a representation showing that MCG18 has homology to two *Caenorhabditis elegans*  
15 proteins.

Figure 22 is a representation showing that MCG18 has homology to a *Saccharomyces pombe* protein.

20 Figure 23 is a representation showing homology of MCG18 to a *Drosophila virilis* protein.

Figure 24 is a representation showing homology of MCG18 to human DnaJ proteins HDJ-2/HSDJ, HDJ-1/HSP40 and HSJ1.

25 Figure 25 is a representation of the nucleotide and corresponding amino acid sequence of murine *mcg18*.

Figure 26 is a representation of homology between human and murine MCG18.

30 Figure 27 depicts nucleotide sequences corresponding to the 5' untranslated region of human *mcg18*.

**Figure 28** depicts a Northern blot showing expression of *mcg18* transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain 15 $\mu$ g RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

**DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

The present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having  
5 homology to a regulator of gene expression or a derivative of said gene regulator.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an  $(\text{HC}_3)_2$   
10 type.

Still more particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- 15 (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C  
20 to the nucleotide sequence set forth in (i), (ii) or (iii).

The present invention also provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.  
25

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- 30 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;

- 16 -

- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

5

Another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock-binding protein or a derivative thereof.

10

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 15 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

20

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1%  
25 v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M  
30 to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least

- 17 -

about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

The present invention extends to nucleic acid molecules with percentage similarities of approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

15 The nucleic acid molecule of the present invention defined by SEQ ID NO:2 is hereinafter referred to as constituting the "*mcg4*" gene. The protein encoded by *mcg4* is referred to herein as "MCG4" and has an amino acid sequence set forth in SEQ ID NO:3. The *mcg4* gene is proposed to encode, in accordance with the present invention, a regulator of gene expression and comprises a novel zinc finger domain, (HC<sub>3</sub>)<sub>2</sub>. A regulator of gene expression includes a transcription factor. Regulation may be at the level of nucleic acid:protein or protein:protein interaction.

The nucleic acid molecule of the present invention defined by SEQ ID NO:4 or 6 is hereinafter referred to as constituting the "*mcg7*" gene. The protein encoded by *mcg7* is referred to herein as "MCG7" and has an amino acid sequence set forth in SEQ ID NO:5 or 7 and is involved in signal transduction. The difference in the nucleotide and amino acid sequence is due to the presence or absence of an exon at nucleotides 183-288.

The nucleic acid molecule of the present invention defined by SEQ ID NO:8 is hereinafter referred to as constituting the "*mcg18*" gene. The protein encoded by *mcg18* is referred to herein as "MCG18" and comprises the amino acid set forth in SEQ ID NO:9.

- 18 -

The present invention extends to the naturally occurring genomic *mcg4*, *mcg7* and *mcg18* nucleotide sequences or corresponding cDNA sequences or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG4, MCG7 or MCG8 or the corresponding genetic sequences. Derivatives  
5 also include single or multiple amino acid substitutions, deletions and/or additions to MCG4, MCG7 or MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to *mcg4*, *mcg7* or *mcg18*. "Additions" to the amino acid or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG4" or "*mcg4*", "MCG7" or "*mcg7*" or "MCG8" or "*mcg18*" includes reference to  
10 all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG4, MCG7 or MCG18.

The *mcg4*, *mcg7* and *mcg18* of the present invention are particularly exemplified herein from humans and in particular from human chromosome 11q13.

15

The present invention extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), reptiles, birds (eg. chickens, ducks, geese, parrots), insects, nematodes, eukaryotic microorganisms and captive wild animals (eg.  
20 deer, foxes, kangaroos). Reference herein to *mcg4* and *mcg18* or their respective proteins MCG4, MCG7 and MCG18 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic  
25 acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic  
30 molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or

- 19 -

both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

- 5 Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg4* gene portion, which *mcg4* gene portion is capable of encoding an MCG4 polypeptide or a functional or immunologically interactive derivative thereof.
- 10 Preferably, the *mcg4* gene portion of the genetic construct is operably linked to a promoter in the vector such that said promoter is capable of directing expression of said *mcg4* gene portion in an appropriate cell.

In addition, the *mcg4* gene portion of the genetic construct may comprise all or part of the gene  
15 fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

20

It is proposed in accordance with the present invention that MCG4 is a transcription factor involved in gene regulation. Mutations in *mcg4* may result in aberrations in gene regulation leading to the development of or a propensity to develop various types of cancer. In this regard, although not wishing to limit the present invention to any one hypothesis or mode of action, it  
25 is proposed that *mcg4* or its expression product may be involved in the tissue-specific or temporal regulation of particular genes.

A deletion or aberration in the *mcg4* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a  
30 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may

be determined by assaying for aberrations in the parents and/or proband of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting  
5 a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

10

Another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg7* gene portion, which *mcg7* gene portion is capable of encoding an *mcg7* polypeptide or a functional or immunologically interactive derivative thereof.

15

Preferably, the *mcg7* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg7* gene portion in an appropriate cell.

20 In addition, the *mcg7* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells  
25 comprising same.

It is proposed in accordance with the present invention that MCG7 is a GEF involved in signal transduction. Mutations in *mcg7* or MCG7 may result in defective control of cell proliferation leading to the development of or a propensity to develop various types of cancer.

30

A deletion or aberration in the *mcg7* gene may also be important in the detection of cancer or



a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents of a subject under investigation.

5

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide  
10 substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Yet another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human  
15 *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg18* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg18* gene portion  
20 in an appropriate cell.

In addition, the *mcg18* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

25

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG18 is a transcription factor  
30 involved in protein folding, protein complex assembly and transit through subcellular compartments. MCG18 may also have a role in tumour suppression. Thus mutations in *mcg18*

- 22 -

may result in the development of or a propensity to develop various types of cancer.

A deletion or aberration in the *mcg18* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a  
5 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents and/or proband of the subject under investigation.

10 According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other  
aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or  
15 a propensity to develop said condition.

The nucleotide substitutions, additions or deletions may be detected by any convenient means including nucleotide sequencing, restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), oligonucleotide hybridization and single stranded conformation  
20 polymorphism analysis (SSCP) amongst many others. An aberration includes modification to existing nucleotides such as to modify glycosylation signal amongst other effects.

In an alternative method, aberrations in the *mcg4*, *mcg7* and *mcg18* genes are detected by screening for mutations in MCG4, MCG7 and MCG18, respectively.

25

A mutation in MCG4, MCG7 or MCG18 may be a single or multiple amino acid substitution, addition and/or deletion. The mutation in *mcg4*, *mcg7* or *mcg18* may also result in either no translation product being produced or a product in truncated form. A mutant may also be an altered glycosylation pattern or the introduction of side chain modifications to amino acid  
30 residues.

- 23 -

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in *mcg4*, *mcg7* or *mcg18* said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4, MCG7 or MCG18 wherein the presence of such a mutation is indicative of or a propensity to  
5 develop said condition.

A particularly convenient means of detecting a mutation in MCG4, MCG7 or MCG18 is by use of antibodies.

- 10 Accordingly another aspect of the present invention is directed to antibodies to MCG4, MCG7 or MCG18 and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG4, MCG7 or MCG18 or may be specifically raised to MCG4, MCG7 or MCG18 or derivatives thereof. In the case of the latter, MCG4, MCG7 or MCG18 or their derivatives may first need to be associated with a carrier molecule.
- 15 The antibodies to MCG4, MCG7 or MCG18 of the present invention are particularly useful as diagnostic agents.

For example, antibodies to MCG4, MCG7 or MCG18 and their derivatives can be used to screen for wild-type MCG4, MCG7 or MCG18 or for mutated MCG4, MCG7 or MCG18 molecules.

- 20 The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG4, MCG7 or MCG18 levels or the presence of wild-type MCG4, MCG7 or MCG18 may be important for diagnosis of certain cancers or a predisposition for development of cancers or for monitoring  
25 certain therapeutic protocols.

- As stated above antibodies to MCG4, MCG7 or MCG18 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to  
30 antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

- 24 -

For example, specific antibodies can be used to screen for wild-type MCG4, MCG7 or MCG18 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for example, as a means for screening for levels of MCG4, MCG7 or MCG18 in a cell extract or other biological fluid or purifying MCG4, MCG7 or MCG18 made by  
5 recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal  
10 or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG4, MCG7 or MCG18 or to a specific mutant phenotype or to a deleted or otherwise altered region.

15

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG4, MCG7 or MCG18 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable  
20 laboratory animal or bird with an effective amount of MCG4, MCG7 or MCG18 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

25

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques  
30 which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting MCG4, MCG7 or MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4, MCG7 or MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4, MCG7 or  
5 MCG18 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

The presence of MCG4, MCG7 or MCG18 may be accomplished in a number of ways such as  
10 by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4,424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

15

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into  
20 contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is  
25 determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the  
30 art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG4, MCG7 or MCG18 including cell extract

or tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the MCG4, MCG7  
5 or MCG18 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding  
10 processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 37°C) to allow binding of any subunit present in the  
15 antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

20 An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-  
25 first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-  
30 bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide

- 27 -

containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled  
5 artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a  
10 fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated,  
15 usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically  
20 coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the  
25 unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

30

As stated above, the present invention extends to genetic constructs capable of encoding MCG4,

MCG7 or MCG18 or functional derivatives thereof. Such genetic constructs are also contemplated to be useful in modulating expression of specific genes in which *mcg4*, *mcg7* or *mcg18* is involved in tissue-specific or temporal regulation.

5 Accordingly, another aspect of the present invention is directed to a genetic construct comprising a nucleotide sequence encoding a peptide, polypeptide or protein and *mcg4*, *mcg7* or *mcg18* or a functional derivative or homologue thereof capable of modulating the expression of said nucleotide sequence.

10 As stated above, MCG18 is proposed to have a role in tumour suppression. Accordingly, it is further proposed in accordance with the present invention to use recombinant MCG18 in pharmaceutical preparations for treating arresting or otherwise ameliorating the effects of certain cancers.

15 Accordingly, another aspect of the present invention contemplates a method for treating, arresting or otherwise ameliorating the effects of a cancer in an animal or bird, said method comprising administering to said animal or bird an effective amount of MCG18 or a functional derivative thereof for a time and under conditions sufficient to treat, arrest or otherwise ameliorate the effects of said cancer.

20

The present invention, therefore, contemplates a pharmaceutical composition comprising MCG18 or a derivative thereof or a modulator of *mcg18* expression or MCG18 activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to hereinafter as the "active ingredients". The active ingredients may also include anti-cancer

25 agents or agents which facilitate actions of MCG18.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be  
30 preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier may be a solvent medium containing, for example, water, ethanol, polyol (for example,



- 29 -

glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin and by the use of surfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, 5 chlorobutanol, phenol, sorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

- 10 Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired 15 ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with 20 the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 25 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions is such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1  $\mu$ g and 2000 mg of active compound:

- 30 The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium

phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of  
5 the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form  
10 should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

The present invention also extends to forms suitable for topical application such as creams,  
15 lotions and gels.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known  
20 in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease  
25 of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the  
30 unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the

treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in  
5 effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as  
hereinbefore disclosed. A unit dosage form can, for example, contain the principal active  
compound in amounts ranging from 0.5  $\mu\text{g}$  to about 2000 mg. Expressed in proportions, the  
active compound is generally present in from about 0.5  $\mu\text{g}$  to about 2000 mg/ml of carrier. In  
the case of compositions containing supplementary active ingredients, the dosages are  
10 determined by reference to the usual dose and manner of administration of the said ingredients.

Effective amounts contemplated by the present invention include those amounts effective to  
ameliorate a condition. For example, it is envisaged that effective amounts would range from  
about 0.001  $\mu\text{g/kg}$  body weight to about 100 mg/kg body weight. Alternatively, effective  
15 amounts of about 0.01  $\mu\text{g/kg}$  body weight to about 10 mg/kg body weight or even 0.1  $\mu\text{g/kg}$   
body weight to about 1 mg/kg body weight. Administration may be per minute, hour, day, week,  
month or year or may only be a once off administration.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable  
20 of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating  
*mcl8* expression or MCG18 activity. The vector may, for example, be a viral vector.

As stated above, the present invention further contemplates a range of derivatives of MCG18.  
Derivatives include fragments, parts, portions, mutants, homologues and analogues of the  
25 MCG18 polypeptide and corresponding genetic sequence. Derivatives also include single or  
multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple  
nucleotide substitutions, deletions and/or additions to the genetic sequence encoding MCG18.  
"Additions" to amino acid sequences or nucleotide sequences include fusions with other  
peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to  
30 "MCG18" includes reference to all derivatives thereof including functional derivatives or MCG18  
immunologically interactive derivatives.

Analogues of MCG18 contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

5

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with  $\text{NaBH}_4$ ; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups  
10 with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with  $\text{NaBH}_4$ .

The guanidine group of arginine residues may be modified by the formation of heterocyclic  
15 condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

20 Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and  
25 other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides.

Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form  
30 a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis  
5 include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 3.

TABLE 3

Non-conventional amino acid	Code	Non-conventional amino acid	Code
5			
$\alpha$ -aminobutyric acid	Abu	L-N-methylalanine	Nmala
$\alpha$ -amino- $\alpha$ -methylbutyrate	Mgab	L-N-methylarginine	Nmarg
aminocyclopropane- carboxylate	Cpro	L-N-methylasparagine	Nmasn
		L-N-methylaspartic acid	Nmasp
10 $\alpha$ -aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbomyl- carboxylate	Norb	L-N-methylglutamine	Nmgln
		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisoleucine	Nmile
15 D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Das	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
20 D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
25 D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
30 D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva

	D-tyrosine	Dtyr	$\alpha$ -methyl-aminoisobutyrate	Maib
	D-valine	Dval	$\alpha$ -methyl- $\gamma$ -aminobutyrate	Mgabv
	D- $\alpha$ -methylalanine	Dmala	$\alpha$ -methylcyclohexylalanine	Mchexa
	D- $\alpha$ -methylarginine	Dmarg	$\alpha$ -methylcyclopentylalanine	Mcpen
5	D- $\alpha$ -methylasparagine	Dmasn	$\alpha$ -methyl- $\alpha$ -naphthylalanine	Manap
	D- $\alpha$ -methylaspartate	Dmasp	$\alpha$ -methylpenicillamine	Mpen
	D- $\alpha$ -methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D- $\alpha$ -methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D- $\alpha$ -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D- $\alpha$ -methylisoleucine	Dmile	N-amino- $\alpha$ -methylbutyrate	Nmaabu
	D- $\alpha$ -methylleucine	Dmleu	$\alpha$ -naphthylalanine	Anap
	D- $\alpha$ -methyllysine	Dmlys	N-benzylglycine	Nphe
	D- $\alpha$ -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D- $\alpha$ -methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D- $\alpha$ -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D- $\alpha$ -methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D- $\alpha$ -methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D- $\alpha$ -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D- $\alpha$ -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D- $\alpha$ -methyltyrosine	Dmtty	N-cyclodecylglycine	Ncdec
	D- $\alpha$ -methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmngln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmnglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl)glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

	D-N-methyllysine	Dnmlys	N-methyl- $\gamma$ -aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmt
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpn
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
5	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyl- $\alpha$ -naphthylalanine	Nmanap
10	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	$\gamma$ -aminobutyric acid	Gabu	N-( <i>p</i> -hydroxyphenyl)glycine	Nhtyr
	L- <i>t</i> -butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L- $\alpha$ -methylalanine	Mala
15	L- $\alpha$ -methylarginine	Marg	L- $\alpha$ -methylasparagine	Masn
	L- $\alpha$ -methylaspartate	Masp	L- $\alpha$ -methyl- <i>t</i> -butylglycine	Mtbug
	L- $\alpha$ -methylcysteine	Mcys	L-methylethylglycine	Metg
	L- $\alpha$ -methylglutamine	Mgln	L- $\alpha$ -methylglutamate	Mglu
	L- $\alpha$ -methylhistidine	Mhis	L- $\alpha$ -methylhomophenylalanine	Mhphe
20	L- $\alpha$ -methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L- $\alpha$ -methylleucine	Mleu	L- $\alpha$ -methyllysine	Mlys
	L- $\alpha$ -methylmethionine	Mmet	L- $\alpha$ -methylnorleucine	Mnle
	L- $\alpha$ -methylnorvaline	Mnva	L- $\alpha$ -methylornithine	Morn
	L- $\alpha$ -methylphenylalanine	Mphe	L- $\alpha$ -methylproline	Mpro
25	L- $\alpha$ -methylserine	Mser	L- $\alpha$ -methylthreonine	Mthr
	L- $\alpha$ -methyltryptophan	Mtrp	L- $\alpha$ -methyltyrosine	Mtyr



- 37 -

L- $\alpha$ -methylvaline	Mval	L-N-methylhomophenylalanine	Nmhph
N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe
1-carboxy-1-(2,2-diphenyl- ethylamino)cyclopropane	Nmbc		

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having  $(CH_2)_n$  spacer groups with  $n=1$  to  $n=6$ ,  
 10 glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of  $C_\alpha$  and  $N_\epsilon$ -methylamino acids, introduction of double bonds between  $C_\alpha$  and  $C_\beta$  atoms of amino acids and  
 15 the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

Such analogues also apply in respect of MCG4 and MCG7.

20

The present invention further contemplates chemical analogues of MCG18 capable of acting as antagonists or agonists of MCG18 or which can act as functional analogues of MCG18. Chemical analogues may not necessarily be derived from MCG18 but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to  
 25 mimic certain physiochemical properties of MCG18. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

The identification of MCG18 permits the generation of a range of therapeutic molecules capable of modulating expression of MCG18 or modulating the activity of MCG18. Modulators  
 30 contemplated by the present invention includes agonists and antagonists of MCG18 expression. Antagonists of MCG18 expression include antisense molecules, ribozymes and co-suppression

molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of MCG18 include molecules which overcome any negative regulatory mechanism. Antagonists of MCG18 include antibodies and inhibitor peptide fragments.

5

These types of modifications may be important to stabilise MCG18 if administered to an individual or for use as a diagnostic reagent.

Other derivatives contemplated by the present invention include a range of glycosylation variants  
10 from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

Another embodiment of the present invention contemplates a method for modulating expression  
15 of MCG18 in a human, said method comprising contacting the *mcg18* gene encoding MCG18 with an effective amount of a modulator of *mcg18* expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression of *mcg18*. For example, a nucleic acid molecule encoding MCG18 or a derivative thereof may be introduced into a cell to facilitate protection of that cell from becoming cancerous.

20

Another aspect of the present invention contemplates a method of modulating activity of MCG18 in a human, said method comprising administering to said mammal a modulating effective amount of a molecule for a time and under conditions sufficient to increase or decrease MCG18 activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative  
25 of MCG18 or a chemical analogue or truncation mutant of MCG18.

The present invention is further described with reference to the following non-limiting Examples.

- 39 -

**EXAMPLE 1**

A human gene (designated *mcg4*) was identified on chromosome 11q13 that on the basis of sequence homology is predicted to encode a putative transcription factor of 310 amino acids  
5 (Fig. 1). *mcg4* is transcribed in several different cell lines (Fig. 7).

**EXAMPLE 2**

The expressed sequence tag (EST) database contains partial sequence data for the murine (Fig.  
10 2) and nematode (Fig. 3) homologues of *mcg4*.

**EXAMPLE 3**

MCG4 contains a sequence of cysteine residues within the N-terminal region of the protein that  
15 resembles zinc-finger binding domains of a novel type, ie.  $(HC_3)_2$  [Fig. 4].

**EXAMPLE 4**

Sensitive sequence homology searches reveal that related cysteine-containing motifs are present  
20 in another *C. elegans* protein (Fig. 5) as well as the GATA-binding transcription factor from *S. pombe* (Fig. 6).

**EXAMPLE 5**

25 *mcg4* will have commercial value due to its likelihood of encoding a novel transcription factor that is highly conserved amongst organisms, thus suggesting an integral role in gene regulation. *mcg4* may also be involved in some way in tissue-specific or temporal regulation of certain genes, thus making it a potential target for modulating expression of those downstream effectors.

### EXAMPLE 6

Nucleotide sequence data generated from cosmid clone cSRL-72c4 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) was aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul *et al* 1990) and was found to match numerous human and mouse entries (Table 4 and Figure 2). These matching ESTs were further used to identify overlapping entries in the EST database (Table 5). The nucleotide sequences of these human ESTs were compiled using MacVector 4.2.1 software (IBI-Kodak) to produce the cDNA sequence shown in Figure 1. EST entries 10 AA074703 and AA134788 are closely related at the nucleotide level to *mcg4* and it is, therefore, likely that *mcg4* is a member of a newly discovered gene family (Figure 8).

The cDNA sequence of *mcg4* was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al*, 1990) at 15 the National Center for Biotechnology Information (<http://www.ncbi.nih.gov.nlm>). As the protein appeared to be novel, a translation of the longest reading frame for the *mcg4* cDNA was aligned to the EST database using the program TBLASTN, which performed a dynamic translation of the EST database in all 6 frames. The search results indicated that the nematode *C. elegans* had an MCG4-like protein (Figure 3), with the matching domains containing a spatial 20 sequence of Cysteine and Histidine residues which resembled a zinc-finger structure (Figure 4). The program BLASTP was used, therefore, to conduct sensitive searches of the protein databases for similar zinc-finger motifs. A weak match to the putative zinc-finger domain was observed for another protein from *C. elegans* (Figure 5) and a poorer match for the GATA-binding transcription factor from *S. pombe* (Figure 6). The putative initiation codon of human 25 *mcg4* is not preceded by an in-frame stop codon and it is therefore possible that the cDNA described in Figure 1 is a truncated form. However, sequence alignment of human and mouse *mcg4* ESTs showed a lower degree of nucleotide conservation prior to the assigned initiation codon, thus supporting the notion that the region represents the 5' UTR (Figure 9). To determine the expression pattern of *mcg4*, 15µg of the total cellular RNA (RNeasy Mini Kit, 30 Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% w/v MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer

using 20 x SSC (Sambrook *et al*, 1989). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (<sup>32</sup>P-dCTP) cDNA probe (Church and Gilbert, 1984) for *mcg4*. After washes in 0.1 x SSC/0.1% w/v SDS at 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg4* is expressed as a 1.6kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 7).

### EXAMPLE 7

A human gene (designated *mcg7*) was identified and isolated from chromosome 11q13 which encodes a protein that bears striking homology with guanine nucleotide exchange factors (GEFs) from a wide variety of organisms (Fig. 12).

### EXAMPLE 8

The composite *mcg7* cDNA sequence is at least 2.4kb in length and Figure 13(a) shows a predicted translation product of at least 609 amino acids beginning at methionine 120. An alternative start site due to alternate exon splicing (indicated in lower case) may yield a protein of 671 amino acids starting at methionine 58 (Fig. 13a).

### EXAMPLE 9

An *mcg7* homologue from *C. elegans* has been identified, the product of which is highly conserved with that of MCG7 (Fig. 14). There are several salient features of the protein which have been underlined in Fig. 14 - namely: a guanine nucleotide binding region, a diacylglycerol binding region, and "EF-hand"-calcium binding regions. In addition, there are several potential cAMP, protein kinase C, and casein kinase II phosphorylation sites, as well as a number of potential sites for glycosylation (not indicated).

### EXAMPLE 10

A number of partial human and murine EST clones exist for *mcg7*. The GenBank database

contains a cDNA (Acc. no. Y12336) encoding a full-length open reading frame (ORF) for human *mcg7* as well as a partial murine *mcg7* ORF (Y12339). In addition, the complete genomic sequence of the human *mcg7* gene is contained within GenBank entry AC000134.

5

**EXAMPLE 11**

The best characterised GEFs are members of the family of *ras* oncoproteins, which play a pivotal role in signal transduction and when mutated are responsible for tumour development. A variety of therapeutic regimes for cancer treatment have been designed to specifically interfere with the *ras* signalling pathways. There is potential, therefore that the product of *mcg7* could also be a target for such clinical strategies.

**EXAMPLE 12**

15 The nucleotide sequence for *mcg7* cDNA was extended 5' with genomic DNA sequence from Genbank accession number AC000134 (positions 1-321) and analysed for additional coding sequence 5' to the putative initiation codon (nt 681-683) (Fig. 16). An additional in-frame ATG occurs at position nt 495-497 when the alternatively splice exon (position nt 504-609) is present (also shown in Fig. 13(a)). This closely matches the Kozak consensus. When this exon is  
20 absent, then the ATG is not in-frame and other possible initiation codons are absent (resulting translation shown in lower case lettering) (also shown in Fig. 13(b)). Further evidence that the initiation codon at position nt 681-683 is the true initiation site is given in Figure 15.

Alignment of human and a partial murine *mcg7* cDNA sequences is shown in Figure 15. The  
25 putative initiation codon is at position nt 360-362. Both murine ESTs appear to have an upstream in-frame stop codon at position nt 326-328, downstream of the differentially spliced exon and the sequence alignment thus suggests that this region represents the 5' UTR of *mcg7*.

Furthermore, similarity with the *C. elegans* homologue strongly suggest that the ATG codon at  
30 position nt 360-362 encodes the N-terminus of MCG7.

### EXAMPLE 13

Figure 17 shows data from experiments indicating that a truncated version of MCG7 when expressed as a GST fusion protein (construct B in Fig. 18) can function as a Ras-guanine nucleotide exchange factor. In brief, Ras (unprocessed and as a GST fusion protein) is loaded with <sup>3</sup>H-GDP then incubated in the presence of excess cold GTP ± GST-MCG7. Full details of this assay can be found in Porfiri *et al.*

### EXAMPLE 14

10

Nucleotide sequence data generated from cosmid clone cSRL-20h12 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) were aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul *et al.*, 1990) and was found to match GenBank entries T78563 (clone 113434) TO9103 (clone HIBBP12) and AA035643 (clone 471819). EST clones 113434 and 471819 were obtained from Genome Systems Inc. and these DNAs were sequenced on both strands with gene-specific primers (Table 5) to generate the cDNA sequence of *mcg7* shown in Figures 13(a) and (b).

The cDNA sequence of *mcg7* was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al.*, 1990) and the coding region was assigned on the basis of showing homology to the *C. elegans* protein F25B3.3 (Figure 14). The *mcg7* cDNA composite was suspected to contain a single nucleotide error that originated from clone 471819 and the correct nucleotide sequence was, therefore, sought by reverse transcription-polymerase chain reaction (RT-PCR) of the cDNA fragment from a human cDNA pool. Total RNA was extracted from a human lymphoblastoid cell line using an RNeasy Mini Kit (Qiagen). cDNA synthesis was conducted with the reverse transcriptase Superscript II RNaseH- (GIBCO, BRL) and random hexamers using the procedure recommended by the manufacturer (GIBCO, BRL). One fortieth of the cDNA mix was subjected to 35 cycles of PCR using the following cycling conditions: 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 90 seconds. The 50μl reaction mix consisted of 1x reaction buffer (Dade Scientific), 2mM dNTP mix, 20pmol of primers (see Table 6) MCG7UF (within the

variably spliced exon of Figure 13(b), between nucleotide positions 184-201) and SGCADRV2 (between nucleotide positions 866-846 of Figure 13(a)) and 10 units of Dynazyme (Dade Scientific). The resulting PCR product was cloned into the pGEM-T vector (Promega) using standard methodology and sequenced using gene-specific primers. The correct nucleotide sequence of *mcg7* (as shown in Figure 13(a)) matches that of the recently release GenBank entry Y12336. A partial mouse *mcg7* cDNA sequence can also be found in GenBank entry Y12339.

### EXAMPLE 15

The coding sequence of *mcg7* was cloned into vectors for expression in both bacterial and mammalian cells. In addition to the full-length constructs, the deletion constructs shown in Figure 18 were designed to retain the guanine nucleotide exchange (GEF) domain. For prokaryotic expression, the *mcg7* coding region was inserted downstream of and in-frame with the Sj26 cassette of the pGEX (Pharmacia) series of vectors (Smith and Johnson, 1988) using standard cloning techniques (Sambrook *et al*, 1989). For mammalian expression, the *mcg7* coding sequence was first *myc*-tagged at the N-terminus and then ligated into the expression vector pc Exv-n using standard cloning techniques. Ligation junctions of the constructs were sequences as the cloning strategies inadvertently changed or introduced additional amino acids as shown below.

20

**Construct (A):** EST clone 113434 was digested with *ApaI* (Figure 13(a), nucleotide positions 1022 to >2416 (within the vector)), blunt-ended with T4 DNA polymerase according to the specifications of the manufacturer (New England Biolab) and ligated into the *SmaI* site of pGEX-3X.

25

Sequence of the pGEX and *mcg7* (underlined) junction:

pGEX-3X                      *mcg7* (1022)  
Sj26 ... GGG ATC CCC CTG GTC [SEQ ID NO:19]

additional amino acids    Gly   Ile   Pro

30

**Construct (B):** EST clone 113434 was digested with *EcoRI* (Figure 13(a), nucleotide



pGEX-1 *mcg7* (695)

additional amino acids    Glu   Phe   Gly   Thr   Ser

15

-----myc-tag-----	vector	BamHI	mcg7 5' UTR (337)	start
ATGGAGCAGAAGCTGATCTCCGAGGAGGACCTG	CCCCGGGCAGCT	ggatccG	<u>CAGCCCCACCCCGCGCGCGGCATG</u>	
M E Q K L I S E E D L	P G A A G S	A A H P A P A A M		
	-----additional amino acids-----			

Sequence of the pGEX and *mcg7* [SEQ ID NO:23/24] (underlined) junction:

- 46 -

pGEX-2      *Bam*HI *mcg7* (337)  
 Sj26    ... gga tcc GCA GCC CAC CCC GCG CCG GCG GCC ATG  
          Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met  
          -----additional amino acids-----

5

**EXAMPLE 16**

Overnight bacterial cultures containing the pGEX plasmid were used to inoculate 500ml of Luria Broth media containing 50µg/ml ampicillin. The cultures were grown to an OD of ~0.8 and then  
 10 induced with 1mM of IPTG for up to 3 hours at 37°C. The bacteria were pelleted and resuspended in 15 ml of STE buffer (10mM Tris pH 8.0, 150 mM NaCl and 1mM EDTA) with 1 mg/ml lysozyme. The mixture was left on ice for more than 1 hour and subsequent steps were performed at 4°C. Protease inhibitors aprotinin, pepstatin and leupeptin were added at final concentrations of 25µg/ml, prior to the addition of Triton-X-100 (2% v/v final) and n-lauroyl  
 15 sarcosine (1.5% w/v final). The lysate was sonicated for ~1 minute and pelleted at 14,000 x g for 15 minutes. 100 µl of 50% w/v glutathione-sephadex bead slurry (in PBS) was added per ml of supernatant. Following a 30 minute incubation at 4°C, the beads were washed three times with NETN (20mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5% NP40), once with NETN-HS (equivalent to NETN but with 1M NaCl), and once in NETN. The bound protein  
 20 was directly analysed by SDS-polyacrylamide gel electrophoresis (PAGE) as described below or the bound protein was eluted from the beads with the following elution buffer (50mM Tris pH 8.0, 150mM NaCl, 5mM MgCl<sub>2</sub>, 1mM DTT, 10mM reduced glutathione) for use in GDP release assays.

25

**EXAMPLE 17**

Twenty microlitres of GST-sepharose-bound MCG7 were added to an equal volume of 2 x  
 30 sample loading dye (100mM Tris pH6.8, 2% v/v mercaptoethanol, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol), boiled for 5 min and loaded onto a 7.5% w/v SDS-PAGE gel (Sambrook *et al*, 1989). The Coomassie brilliant blue stained gel (Sambrook *et al*, 1989)

typically displayed a protein doublet, running between 87-95 kDa consisting of the MCG7-GST fusion and a slightly smaller, co-purified contaminating *E. coli* protein of ~105kDa. The calculated molecular weight of full-length MCG7 is 77.5 kDa (Construct (D)) and the GST component has a molecular weight of 26kDa, hence, the recombinant protein runs slightly smaller than predicted. A Western blot of the same gel probed with anti-GST antibody yields an MCG7-specific band at the same position as that of the stained gel.

### EXAMPLE 18

10 Assumptions: (a) GST-Ras molecular weight = 50 kD; (b) Concentration of GST-Ras solution = 1mg/ml = 20 $\mu$ M; (c) [<sup>3</sup>H]-GDP is 1mCi/ml and 13.3Ci/mmol, therefore [ <sup>3</sup>H]-GDP concentration = 75  $\mu$ M and 1pmol [<sup>3</sup>H]-GDP=15,466 cpm; (d) Elution buffer = Buffer E = 20 mM Tris-Cl, pH7.5; 50mM NaCl; 5mM MgCl<sub>2</sub>; 1mM DTT (added just before use). Buffer E + BSA= Buffer E+1mg/ml BSA (added just before use).

15

Mix together, in the following order and mix well after each addition:

10 $\mu$ l (=10 $\mu$ g) GST-Ras (@1mg/ml in Buffer E), 463 $\mu$ l Buffer E + BSA, 7 $\mu$ l [<sup>3</sup>H]-GDP, 10ml 490  $\mu$ M EDTA. Incubate @ RT for 10 min. Add 10 $\mu$ l 0.5 M MgCl<sub>2</sub> and mix well. Incubate @ RT for 10 min. Place on ice. During the first incubation the excess EDTA concentration is 5mM, during the second incubation the excess Mg concentration is 5mM. The [<sup>3</sup>H]-GDP concentration is 1 $\mu$ M and the final concentration of GST-Ras is 400nM. Thus 20ml of the final mix will contain 8pmol of GST-Ras protein. Specific activity of GDP is 15,446 cpm/pmol x (1/1.4) = 11,047 cpm/pmol.

25

### EXAMPLE 19

Exchange Ras with labelled GDP as above. Add unlabelled GTP (stock = 100mM, pH7) to 1 mM. Adjust Mg concentration by adding 5 $\mu$ l 0.5 EDTA to labelled Ras, 5 $\mu$ l 0.5M EDTA to 500 $\mu$ l MCG7, and 5 $\mu$ l 0.5M EDTA to 500 $\mu$ l Buffer E + BSA. On ice set up microfuge tubes with 40 $\mu$ l Ras-GDP (in triplicate) with 40 $\mu$ l MCG7 or Buffer E + BSA (control). Transfer tubes to heat block @ 25°C and incubate for 10, 20 or 30 min. Stop exchange reactions with 1ml of

- 48 -

ice cold buffer E and place on ice. Pre-soak nitrocellulose filters, pore size 45 $\mu$ m, in Buffer E. Assemble the vacuum manifold apparatus (Millipore) with wet filters and plug the wells with rubber bunds. Switch on the vacuum pump. Remove the first plug, aliquot the sample and once it has been sucked through, wash the filter with 10ml of ice cold Buffer E. Remove next plug  
5 etc and continue round the manifold. Take manifold apart. Pin the filters to a pin board reserved for [ $^3$ H]. Air dry. Take up in 4ml scintillation fluid and count. These studies have been carried out with a truncated MCG7-GST fusion protein (amino acids 341 of Figure 13a to stop encoded within construct B).

10

**EXAMPLE 20**

A human gene was identified from chromosome 11q13 that encodes a new member of the DnaJ family of proteins (designated MCG18). This gene (*mcg18*) is expressed as an ~1.4kb mRNA (Fig. 28) and is predicted to encode a 241 amino acid product (Fig. 19).

15

**EXAMPLE 21**

MCG18 has partial homology to *E. coli* dnaJ and other human DnaJ family members in that it contains the J domain (Fig. 20).

20

**EXAMPLE 22**

MCG18 has greatest homology to functionally undefined proteins from *C. elegans* (Fig. 21) and *S. pombe* (Fig. 22) that also feature the J domain but maintain sequence similarity through the  
25 central and C-terminal regions of the proteins.

**EXAMPLE 23**

The J domain is proposed to mediate interaction with heat shock protein (Hsp70) 70 and consist  
30 of some 70 amino acids, frequently located at the N-terminus of the protein. One of these proteins, tumorous imaginal discs (Tid58) from *Drosophila virilis* (Fig. 23) functions as a

tumour suppressor.

#### EXAMPLE 24

- 5 A comparison of homology between MCG18 and human DnaJ proteins HDJ-2/H5DJ, HDJ-1/HSP40 and HSJ1 is shown in Fig. 24.

#### EXAMPLE 25

- 10 During the sequence characterisation of the *VRF/VEGFB* promoter region on cosmid CLGW4 [Grimmond *et al*, 1996], which maps to chromosome 11q13 the inventors identified a sequence that exactly matched numerous human and mouse expressed sequence tags (ESTs) in the EST database from a gene which we designated *mcg18*. EST clones for human (GenBank accession number T69741, clone 108172; accession number H40901, clone 177008) and mouse *mcg18*  
15 (accession number W34884, clone 350966; accession number W64183, clone 385535) were obtained from Genome Systems Inc. and sequenced with the gene-specific primers shown in Table 7. The EST clones listed in Table 8 were also utilised in generating the full-length coding sequence for human (Figure 19) and mouse (Figure 25) *mcg18*. The EST database also contained *mcg18* cDNA entries that were alternately (or partially) spliced, and in order to  
20 understand their ability to encode new polypeptides, the gene structure of *mcg18* was determined by sequencing human and mouse genomic templates with gene-specific primers.

- Genomic fragments containing the human [Grimmond *et al*, 1996] and murine genes [Townson *et al*, 1996] have been previously reported. Cosmid CLGW4 contains the entire human gene  
25 and  $\lambda$ 121 contains the entire mouse gene, as determined by direct sequencing of the templates with the oligonucleotides listed in Table 7. Plasmids containing sub-fragments of  $\lambda$ 121 and cosmid CLGW4 were prepared using plasmid purification kits (Qiagen) and sequenced as described previously [Grimmond *et al*, 1996; Townson *et al*, 1996] using primers designed against cDNA and genomic sequences. The BLAST suite of programs [Altschul *et al*, 1990]  
30 was used to compare the sequence data against the nucleotide and protein databases at the National Center for Biotechnology Information (<http://www.ncbi.nih.gov.nlm>). The sequence

data were compiled using MacVector 4.2.1 software (IBI-Kodak). ClustalW sequence alignments [Thompson *et al*, 1994] were conducted using the Australian National Genome Information Service computer faculty at the University of Sydney, Australia.

5 The cDNA sequence of human *mcg18* (Figure 19) was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX [Altschul *et al*, 1990] and the coding region was identified on the basis of showing homology to the DnaJ family of proteins (Figure 20). The DnaJ domain is encoded within the longest open reading frame and the assigned initiation codon is preceded by an in-frame stop codon (Figure  
10 27). Similar database search results were obtained for the mouse *mcg18* cDNA, and the alignment of human and mouse protein sequences is shown in Figure 26. MCG18 has greatest homology to gene products from *C. elegans* (Figure 21) and *S. pombe* (Figure 22). Although it shares a similar J-domain, MCG18 does not contain other domains described for the tumour suppressor gene from *D. virilis* (Figure 23), nor is it a homologue of other reported human J-  
15 domain-containing proteins (Figure 24).

To determine the expression pattern of *mcg18*, 15µg of total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer  
20 using 20 x SSC (Sambrook *et al*, 1986). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (<sup>32</sup>P-dCTP) cDNA probe (Church and Gilbert, 1984) for *mcg18*. After washes in 0.1 x SSC/0.1% w/v SDS for 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg18* is expressed as a 1.4kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 28).

- 51 -

TABLE 4

ESTs matching *mcg4*

accession number	seq. run	organism	score	E value	N
gb AA399110 AA399110	zt89e06.s1	Soares testis NHT Homo sa...	1136	4.0e-168	2
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone 2...	1521	5.3e-168	4
gb AA514406 AA514406	nf57d01.s1	NCI_CGAP_Co3 Homo sapiens...	931	5.5e-166	3
gb AA544946 AA544946	vk38e02.r1	Soares mouse mammary glan...	1207	8.4e-164	2
gb AA450076 AA450076	zx42a04.s1	Soares total fetus Nb2HF8...	691	2.3e-160	4
gb AA535731 AA535731	nf88f07.s1	NCI_CGAP_Co3 Homo sapiens...	796	3.5e-158	4
gb W79710 W79710	zd86f01.r1	Soares fetal heart NbHH19...	1644	1.1e-157	4
gb AA503531 AA503531	ne47e08.s1	NCI_CGAP_Co3 Homo sapiens...	736	4.0e-156	4
gb AA450132 AA450132	zx42a04.r1	Soares total fetus Nb2HF8...	1955	3.9e-155	1
gb AA398068 AA398068	zt89f06.r1	Soares testis NHT Homo sa...	1315	5.4e-148	2
gb W60405 W60405	zd29h08.r1	Soares fetal heart NbHH19...	1022	1.8e-139	4
gb W81382 W81382	zd86f01.s1	Soares fetal heart NbHH19...	605	3.5e-125	5
gb AA047617 AA047617	zf13f07.s1	Soares fetal heart NbHH19...	922	4.6e-125	2
gb AA282175 AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapien...	1577	2.0e-123	1
gb AA242159 AA242159	my30d04.r1	Barstead mouse pooled org...	866	7.7e-117	2
gb AA068680 AA068680	mm61a05.r1	Stratagene mouse embryoni...	1280	1.6e-98	1
gb W46766 W46766	zc36b07.s1	Soares senescent fibrobla...	506	9.6e-92	3
gb N93704 N93704	zb51c04.s1	Soares fetal lung NbHL19W...	584	9.0e-91	4
gb AA155210 AA155210	mr98e01.r1	Stratagene mouse embryoni...	840	7.6e-87	2
gb AA366022 AA366022	EST76915	Pineal gland II Homo sapien...	1077	2.4e-81	1
gb AA037691 AA037691	zk34h12.s1	Soares pregnant uterus Nb...	949	2.1e-80	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor ...	1016	3.1e-76	1
dbj C00696 C00696	HUMGS0008251	Human Gene Signature, ...	1009	1.2e-75	1
gb T98249 T98249	ye59a07.s1	Homo sapiens cDNA clone 1...	998	6.7e-75	1
gb W21588 W21588	zb51c04.r1	Soares fetal lung NbHL19W...	484	1.1e-69	4
gb H32171 H32171	EST107015	Rattus sp. cDNA 5' end.	828	1.1e-60	1
gb AA108092 AA108092	mm89e06.r1	Stratagene mouse embryoni...	782	1.3e-60	2
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4Nb...	665	2.5e-60	2
gb AA037690 AA037690	zk34h12.r1	Soares pregnant uterus Nb...	540	9.4e-53	2
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapien...	535	5.4e-48	2
gb N46760 N46760	yy51g06.r1	Homo sapiens cDNA clone 2...	665	9.5e-47	1
gb W23584 W23584	zc71d03.s1	Soares fetal heart NbHH19...	457	1.8e-44	2
gb W42214 W42214	mc69h09.r1	Soares mouse embryo NbME1...	460	1.3e-38	3
gb AA244877 AA244877	mx25a04.r1	Soares mouse NML Mus musc...	429	2.9e-25	1
gb W32939 W32939	zc07h03.r1	Soares parathyroid tumor ...	320	4.8e-18	1

- 52 -

**TABLE 5**  
**ESTs matching AA074703 (*mcg4*-related cDNA)**

Database: Non-redundant Database of GenBank EST Division  
 1,222,625 sequences; 449,352,662 total letters.

			Smallest Sum			High Probability		
Sequences producing High-scoring Segment Pairs:			Score	P(N)	N	Score	P(N)	N
accession number	seq. run	organism	score	E value	N			
gb AA074703 AA074703	zm76g07.r1	Stratagene neuroepitheli...	2071	4.0e-167	1			
gb AA068680 AA068680	mm61a05.r1	Stratagene mouse embryon...	1270	4.4e-145	4			
gb AA134788 AA134788	zm81g02.r1	Stratagene neuroepitheli...	946	1.3e-144	5			
gb AA399110 AA399110	zt89e06.s1	Soares testis NHT Homo s...	520	8.7e-119	6			
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone ...	582	9.6e-110	7			
gb AA282175 AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapie...	771	9.4e-80	3			
gb W81382 W81382	zd86f01.s1	Soares fetal heart NbHH1...	329	1.6e-75	6			
gb AA544946 AA544946	vk38e02.r1	Soares mouse mammary gla...	644	9.6e-63	2			
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor...	294	4.5e-42	4			
gb W57106 W57106	md57c12.r1	Soares mouse embryo NbME...	394	1.9e-30	2			
gb AA244877 AA244877	mx25a04.r1	Soares mouse NML Mus mus...	162	2.1e-27	4			
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4N...	230	3.7e-23	3			
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapie...	139	2.3e-19	3			
gb H32171 H32171	EST107015	Rattus sp. cDNA 5' end.	207	2.6e-10	2			
gb W79710 W79710	zd86f01.r1	Soares fetal heart NbHH1...	157	0.0073	1			



**TABLE 6**  
***mcg7*-specific oligonucleotides**

5	name	sequence (5' to 3')	SEQ ID NOs.
	M1044R	GGA CAA AGT GTG TGA TGA ACC	SEQ ID NO:25
	MCG7-GEF-REV2	CTC ATC CTC CGT CTG ATA CTG	SEQ ID NO:26
	M7R	GTA GAT GTG GAT CAG CTT GG	SEQ ID NO:27
	MCG7 CA FOR	AGG TGG AGA ATG GTC AAGG	SEQ ID NO:28
10	MCG7-GEF-REV	GTC ATA GTC TGT CTC CTA CT	SEQ ID NO:29
	MCG7 GEF FOR	ACA TAG ACA GCG TGC CTA CC	SEQ ID NO:30
	MCG7-PKC-REV	TAC AAC CTT AGG GAC ACC AG	SEQ ID NO:31
	MCG7-PKC-FOR	TGC TGA GCC TGC TCA CGG TG	SEQ ID NO:32
	T09103F	CAA GTG AAC AGC ACG TCC	SEQ ID NO:33
15	M7F	GAC TAT CTC AAG GAC CAG CTG	SEQ ID NO:34
	MCG7UF	GGT TCG GTC CGA GCC CGG	SEQ ID NO:35
	SGCADRV2	GGA GCG ATA CTC CAA GTA GGT	SEQ ID NO:36

**TABLE 7**  
***mcg18*-SPECIFIC OLIGONUCLEOTIDES**

	<b>name</b>	<b>sequence 5' to 3'</b>
5	HVESTF	AGC GGG CCA GGC CCC TTC [SEQ ID NO:37]
	HV195F	CAT CCT GGT CCA ATG CGC TC [SEQ ID NO:38]
	HV387F2	GCA CTG AGG AAG TTA AAC GAG C [SEQ ID NO:39]
	HV408R	GCT CGT TTA ACT TCC TCA GTG C [SEQ ID NO:40]
	EXON1REV	GCT CAG CTC CAC AAA GCG GCT [SEQ ID NO:41]
10	HVEST426F	ACC AGC TCC GCT CAG GTA G [SEQ ID NO:42]
	HVEST623R	TCC AGG AGC TGT GTG TTT GG [SEQ ID NO:43]
	SGVESTF3	CCA GTT TCA CAG CGT GAG G [SEQ ID NO:44]
	HVEST631R	CAG CAT GAG GAG GAG GCA G [SEQ ID NO:45]

**TABLE 8**  
**EST CLONE SEQUENCES USED TO GENERATE HUMAN AND MOUSE**  
***mcg18* cDNA SEQUENCE COMPOSITES**

<u>EST clone number</u>	<u>organism</u>	<u>GenBank accession number</u>
1g2815	human	D45683
001-T2-18	human	F17225
273748	human	N37043
177008	human	H40901 and H40939
258011	human	N30776
276887	human	N44004
108172	human	T69741
307529	human	W21083 and W32579
342027	human	W60283
354288	mouse	W44038
350966	mouse	W348844
426261	mouse	AA002868
368185	mouse	W53911
385535	mouse	W64183
404472	mouse	W82959
406437	mouse	W83482

**BIBLIOGRAPHY**

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *J. Mol. Biol.* 215: 403-410.
2. Church, G., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* 18: 1991-1995.
3. Porfiri *et al.* *J. Biol. Chem.* 269: 22672-22677 (1994).
4. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, USA.
5. Smith, D.B., and Johnson, K.S. (1988) *Gene* 67: 31-40.
6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) *Nucleic Acids Res.* 22: 4673-4680.
7. Beranger, F., Paterson, H., Powers, S., de Gunzburg, J. and Hancock, J.F. (1994) *Molecular and Cellular Biology* 14: 744-758.
8. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G (1996) *Genome Res.* 6: 124-131.
9. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun.* 220: 922-928.

- 57 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: (OTHER THAN US): The Council of The Queensland Institute of Medical Research  
(US ONLY): HAYWARD Nicholas, SILINS Ginters, GRIMMOND Sean, GARTSIDE Michael and HANCOCK, John

(ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 45

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DAVIES COLLISON CAVE  
(B) STREET: 1 LITTLE COLLINS STREET  
(C) CITY: MELBOURNE  
(D) STATE: VICTORIA  
(E) COUNTRY: AUSTRALIA  
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: PCT INTERNATIONAL  
(B) FILING DATE: 22-MAY-1998  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6973  
(B) FILING DATE: 23-MAY-1997  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6974  
(B) FILING DATE: 23-MAY-1997  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6972  
(B) FILING DATE: 23-MAY-1997

- 58 -

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1459
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1460
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1458
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: HUGHES, DR E JOHN L
- (C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: +61 3 9254 2777
- (B) TELEFAX: +61 3 9254 2770
- (C) TELEX: AA 31787

- 59 -

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Cys Xaa Xaa Cys Xaa Gly Xaa Gly  
 5

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1242 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 30..959

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCAGTAAACA CAGAGACTGG GGATCGATC	ATG GGG CTT TGT AAG TGC CCC AAG	53
	Met Gly Leu Cys Lys Cys Pro Lys	
	1 5	
AGA AAG GTG ACC AAC CTG TTC TGC TTC GAA CAT CGG GTC AAC GTC TGC	101	
Arg Lys Val Thr Asn Leu Phe Cys Phe Glu His Arg Val Asn Val Cys		
10 15 20		
GAG CAC TGC CTG GTA GCC AAT CAC GCC AAG TGC ATC GTC CAG TCC TAC	149	
Glu His Cys Leu Val Ala Asn His Ala Lys Cys Ile Val Gln Ser Tyr		
25 30 35 40		
CTG CAA TGG CTC CAA GAT AGC GAC TAC AAC CCC AAT TGC CGC CTG TGC	197	
Leu Gln Trp Leu Gln Asp Ser Asp Tyr Asn Pro Asn Cys Arg Leu Cys		
45 50 55		
AAC ATA CCC CTG GCC AGC CGA GAG ACG ACC CGC CTT GTC TGC TAT GAT	245	
Asn Ile Pro Leu Ala Ser Arg Glu Thr Thr Arg Leu Val Cys Tyr Asp		
60 65 70		
CTC TTT CAC TGG GCC TGC CTC AAT GAA CGT GCT GCC CAG CTA CCC CGA	293	
Leu Phe His Trp Ala Cys Leu Asn Glu Arg Ala Ala Gln Leu Pro Arg		
75 80 85		
AAC ACG GCA CCT GCC GGC TAT CAG TGC CCC AGC TGC AAT GGC CCC ATC	341	
Asn Thr Ala Pro Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ile		
90 95 100		
TTC CCC CCA ACC AAC CTG GCT GGC CCC GTG GCC TCC GCA CTG AGA GAG	389	
Phe Pro Pro Thr Asn Leu Ala Gly Pro Val Ala Ser Ala Leu Arg Glu		
105 110 115 120		

- 60 -

AAG CTG GCC ACA GTC AAC TGG GCC CGG GCA GGA CTG GGC CTC CCT CTG	437
Lys Leu Ala Thr Val Asn Trp Ala Arg Ala Gly Leu Gly Leu Pro Leu	
125 130 135	
ATC GAT GAG GTG GTG AGC CCA GAG CCC GAG CCC CTC AAC ACG TCT GAC	485
Ile Asp Glu Val Val Ser Pro Glu Pro Glu Pro Leu Asn Thr Ser Asp	
140 145 150	
TTC TCT GAC TGG TCT AGT TTT AAT GCC AGC AGT ACC CCT GGA CCA GAG	533
Phe Ser Asp Trp Ser Ser Phe Asn Ala Ser Ser Thr Pro Gly Pro Glu	
155 160 165	
GAG GTA GAC AGC GCC TCT GCT GCC CCA GCC TTC TAC AGC CGA GCC CCC	581
Glu Val Asp Ser Ala Ser Ala Ala Pro Ala Phe Tyr Ser Arg Ala Pro	
170 175 180	
CGG CCC CCA GCT TCC CCA GGC CGG CCC GAG CAG CAC ACA GTG ATC CAC	629
Arg Pro Pro Ala Ser Pro Gly Arg Pro Glu Gln His Thr Val Ile His	
185 190 195 200	
ATG GGC AAT CCT GAG CCC TTG ACT CAC GCC CCT AGG AAG GTG TAT GAT	677
Met Gly Asn Pro Glu Pro Leu Thr His Ala Pro Arg Lys Val Tyr Asp	
205 210 215	
ACG CGG GAT GAT GAC CGG ACA CCA GGC CTC CAT GGA GAC TGT GAC GAT	725
Thr Arg Asp Asp Asp Arg Thr Pro Gly Leu His Gly Asp Cys Asp Asp	
220 225 230	
GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTG CTA	773
Asp Lys Tyr Arg Arg Arg Pro Ala Leu Gly Trp Leu Ala Arg Leu Leu	
235 240 245	
AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG	821
Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg	
250 255 260	
GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT	869
Ala Gly Leu Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu	
265 270 275 280	
GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC	917
Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn	
285 290 295	
CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA	962
Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser *	
300 305 310	
GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT	1022
AATGGGGAGG CTGAGGGCAC CTCCTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA	1082
CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG	1142
GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC	1202
ATGAAGGAGC TGGCAGGTGG AAATAAACAA CAACTTTATT	1242

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 310 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein



- 61 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Leu Cys Lys Cys Pro Lys Arg Lys Val Thr Asn Leu Phe Cys  
 1 5 10 15  
 Phe Glu His Arg Val Asn Val Cys Glu His Cys Leu Val Ala Asn His  
 20 25 30  
 Ala Lys Cys Ile Val Gln Ser Tyr Leu Gln Trp Leu Gln Asp Ser Asp  
 35 40 45  
 Tyr Asn Pro Asn Cys Arg Leu Cys Asn Ile Pro Leu Ala Ser Arg Glu  
 50 55 60  
 Thr Thr Arg Leu Val Cys Tyr Asp Leu Phe His Trp Ala Cys Leu Asn  
 65 70 75 80  
 Glu Arg Ala Ala Gln Leu Pro Arg Asn Thr Ala Pro Ala Gly Tyr Gln  
 85 90 95  
 Cys Pro Ser Cys Asn Gly Pro Ile Phe Pro Pro Thr Asn Leu Ala Gly  
 100 105 110  
 Pro Val Ala Ser Ala Leu Arg Glu Lys Leu Ala Thr Val Asn Trp Ala  
 115 120 125  
 Arg Ala Gly Leu Gly Leu Pro Leu Ile Asp Glu Val Val Ser Pro Glu  
 130 135 140  
 Pro Glu Pro Leu Asn Thr Ser Asp Phe Ser Asp Trp Ser Ser Phe Asn  
 145 150 155 160  
 Ala Ser Ser Thr Pro Gly Pro Glu Glu Val Asp Ser Ala Ser Ala Ala  
 165 170 175  
 Pro Ala Phe Tyr Ser Arg Ala Pro Arg Pro Pro Ala Ser Pro Gly Arg  
 180 185 190  
 Pro Glu Gln His Thr Val Ile His Met Gly Asn Pro Glu Pro Leu Thr  
 195 200 205  
 His Ala Pro Arg Lys Val Tyr Asp Thr Arg Asp Asp Asp Arg Thr Pro  
 210 215 220  
 Gly Leu His Gly Asp Cys Asp Asp Asp Lys Tyr Arg Arg Arg Pro Ala  
 225 230 235 240  
 Leu Gly Trp Leu Ala Arg Leu Leu Arg Ser Arg Ala Gly Ser Arg Lys  
 245 250 255  
 Arg Pro Leu Thr Leu Leu Gln Arg Ala Gly Leu Leu Leu Leu Leu Gly  
 260 265 270  
 Leu Leu Gly Phe Leu Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg  
 275 280 285  
 Ala Ala Ala Asp Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His  
 290 295 300  
 Ile Arg Val Gly Pro Ser  
 305 310

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- 62 -

- (A) LENGTH: 2415 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..2188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC TCC CCA AAA TAT TCC	47
Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser	
1 5 10 15	
CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG CTG TCC	95
His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser	
20 25 30	
CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG CCT GCA GGA	143
Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly	
35 40 45	
AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA ACG CAG CGC CTG	191
Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu	
50 55 60	
TGT GGC CGC GGG ACT CAA GGC TGG CCT GGC TCA AGT GAA CAG CAC GTC	239
Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val	
65 70 75	
CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT TCT GGG GTG GAC GAG CTG	287
Gln Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu	
80 85 90 95	
GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC	335
Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly	
100 105 110	
CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC CTG GAC	383
Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp	
115 120 125	
AAG GGC TGC ACG GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC	431
Lys Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe	
130 135 140	
GAT GAC TCC GGG AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC	479
Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu	
145 150 155	
ATG ATG CAC CCC TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG	527
Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu	
160 165 170 175	
CTC CAC ATC TAC CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG	575
Leu His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln	
180 185 190	
GTG AAA ACG TGC CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG	623
Val Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala	
195 200 205	

- 63 -

GAG TTT GAC TTG AAC CCG GAG TTG GCT GAG CAG ATC AAG GAG CTG AAG Glu Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys 210 215 220	671
GCT CTG CTA GAC CAA GAA GGG AAC CGA CGG CAC AGC AGC CTA ATC GAC Ala Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp 225 230 235	719
ATA GAC AGC GTC CCT ACC TAC AAG TGG AAG CGG CAG GTG ACT CAG CGG Ile Asp Ser Val Pro Thr Tyr Lys Trp Lys Met Ser Gln Val Thr Gln Arg 240 245 250 255	767
AAC CCT GTG GGA CAG AAA AAG CGC AAG ATG TCC CTG TTG TTT GAC CAC Asn Pro Val Gly Gln Lys Lys Arg Lys Lys Ser Leu Leu Phe Asp His 260 265 270	815
CTG GAG CCC ATG GAG CTG GCG GAG CAT CTC ACC TAC TTG GAG TAT CGC Leu Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg 275 280 285	863
TCC TTC TGC AAG ATC CTG TTT CAG GAC TAT CAC AGT TTC GTG ACT CAT Ser Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His 290 295 300	911
GGC TGC ACT GTG GAC AAC CCC GTC CTG GAG CGG TTC ATC TCC CTC TTC Gly Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe 305 310 315	959
AAC AGC GTC TCA CAG TGG GTG CAG CTC ATG ATC CTC AGC AAA CCC ACA Asn Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr 320 325 330 335	1007
GCC CCG CAG CGG GCC CTG GTC ATC ACA CAC TTT GTC CAC GTG GCG GAG Ala Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu 340 345 350	1055
AAG CTG CTA CAG CTG CAG AAC TTC AAC ACG CTG ATG GCA GTG GTC GGG Lys Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly 355 360 365	1103
GGC CTG AGC CAC AGC TCC ATC TCC CGC CTC AAG GAG ACC CAC AGC CAC Gly Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His 370 375 380	1151
GTT AGC CCT GAG ACC ATC AAG CTC TGG GAG GGT CTC ACG GAA CTA GTG Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val 385 390 395	1199
ACG GCG ACA GGC AAC TAT GGC AAC TAC CGG CGT CGG CTG GCA GCC TGT Thr Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys 400 405 410 415	1247
GTG GGC TTC CGC TTC CCG ATC CTG GGT GTG CAC CTC AAG GAC CTG GTG Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val 420 425 430	1295
GCC CTG CAG CTG GCA CTG CCT GAC TGG CTG GAC CCA GCC CGG ACC CGG Ala Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg 435 440 445	1343
CTC AAC GGG GCC AAG ATG AAG CAG CTC TTT AGC ATC CTG GAG GAG CTG Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu 450 455 460	1391
GCC ATG GTG ACC AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu 465 470 475	1439

CTG AGC CTG CTC ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu 480 485 490 495	1487
CTG TAC CAG CTG TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro 500 505 510	1535
ACC AGC CCC ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu 515 520 525	1583
GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val 530 535 540	1631
GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val 545 550 555	1679
GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly 560 565 570 575	1727
AAC TTC CCT TAC CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp 580 585 590	1775
GGC TGC ATC AGC AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser 595 600 605	1823
TCT GTG TTG GGG GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser 610 615 620	1871
AAC TCC TTG CGC CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG Asn Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu 625 630 635	1919
GGC ATC TAC AAG CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys 640 645 650 655	1967
CAC AAG CAG TGC AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala 660 665 670	2015
CAG AGT GTG AGC CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His 675 680 685	2063
AGC CAC CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG Ser His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg 690 695 700	2111
CGA GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val 705 710 715	2159
GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG TGTTGGATC Glu Asp Gly Val Phe Asp Ile His Leu 720 725	2208
AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA GCAGGGAGCC TGGGGGTGTC	2268
GGGGCAGGAG GCTGGGGATG GGGGTGGGAT ATGAGGGTGG CATGCAGCTG AGGGCAGGGC	2328

- 65 -

CAGGGCTGGT GTCCCTAAGG TTGTACAGAC TCTTGTGAAT ATTTGTATTT TCCAGATGGA 2388  
 ATAAAAAGGC CCGTGTAATT AACCTTC 2415

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 728 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE-DESCRIPTION: SEQ ID NO:5:

Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser His  
 1 5 10 15  
 Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser Pro  
 20 25 30  
 Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly Arg  
 35 40 45  
 Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu Cys  
 50 55 60  
 Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val Gln  
 65 70 75 80  
 Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu Gly  
 85 90 95  
 Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro  
 100 105 110  
 Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys  
 115 120 125  
 Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp  
 130 135 140  
 Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met  
 145 150 155 160  
 Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu  
 165 170 175  
 His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val  
 180 185 190  
 Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu  
 195 200 205  
 Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala  
 210 215 220  
 Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile  
 225 230 235 240  
 Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn  
 245 250 255  
 Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu  
 260 265 270

- 66 -

Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser  
 275 280 285  
 Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His Gly  
 290 295 300  
 Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe Asn  
 305 310 315 320  
 Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr Ala  
 325 330 335  
 Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu Lys  
 340 345 350  
 Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly Gly  
 355 360 365  
 Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His Val  
 370 375 380  
 Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr  
 385 390 395 400  
 Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val  
 405 410 415  
 Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala  
 420 425 430  
 Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu  
 435 440 445  
 Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala  
 450 455 460  
 Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu  
 465 470 475 480  
 Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu  
 485 490 495  
 Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr  
 500 505 510  
 Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu  
 515 520 525  
 Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu  
 530 535 540  
 His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp  
 545 550 555 560  
 Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn  
 565 570 575  
 Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly  
 580 585 590  
 Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser  
 595 600 605  
 Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn  
 610 615 620  
 Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly

- 67 -

625		630		635		640
Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His	645		650		655	
Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln	660		665		670	
Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser	675		680		685	
His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg	690		695		700	
Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu	705		710		715	720
Asp Gly Val Phe Asp Ile His Leu	725					

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2309 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 254..2083

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTGTCTCTAG	60
CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC	120
TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAGC CCCATGGGAA	180
CGGGGTTCGG TCCGAGCCCG GTGGGAGGCT CCCGGAGCGC AGCCTGGGCC CAGCCCACCC	240
CGCGCCGGCG GCC ATG GCA GGC ACC CTG GAC CTG GAC AAG GGC TGC ACG	289
Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr	
1 5 10	
GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC GAT GAC TCC GGG	337
Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly	
15 20 25	
AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC ATG ATG CAC CCC	385
Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro	
30 35 40	
TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG CTC CAC ATC TAC	433
Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr	
45 50 55 60	
CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG GTG AAA ACG TGC	481
Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys	
65 70 75	
CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG GAG TTT GAC TTG	529

- 68 -

His	Leu	Val	Arg	Tyr	Trp	Ile	Ser	Ala	Phe	Pro	Ala	Glu	Phe	Asp	Leu	
			80					85					90			
AAC	CCG	GAG	TTG	GCT	GAG	CAG	ATC	AAG	GAG	CTG	AAG	GCT	CTG	CTA	GAC	577
Asn	Pro	Glu	Leu	Ala	Glu	Gln	Ile	Lys	Glu	Leu	Lys	Ala	Leu	Leu	Asp	
		95					100					105				
CAA	GAA	GGG	AAC	CGA	CGG	CAC	AGC	AGC	CTA	ATC	GAC	ATA	GAC	AGC	GTC	625
Gln	Glu	Gly	Asn	Arg	Arg	His	Ser	Ser	Leu	Ile	Asp	Ile	Asp	Ser	Val	
	110					115					120					
CCT	ACC	TAC	AAG	TGG	AAG	CGG	CAG	GTG	ACT	CAG	CGG	AAC	CCT	GTG	GGA	673
Pro	Thr	Tyr	Lys	Trp	Lys	Arg	Gln	Val	Thr	Gln	Arg	Asn	Pro	Val	Gly	
125				130					135						140	
CAG	AAA	AAG	CGC	AAG	ATG	TCC	CTG	TTG	TTT	GAC	CAC	CTG	GAG	CCC	ATG	721
Gln	Lys	Lys	Arg	Lys	Met	Ser	Leu	Leu	Phe	Asp	His	Leu	Glu	Pro	Met	
			145						150					155		
GAG	CTG	GCG	GAG	CAT	CTC	ACC	TAC	TTG	GAG	TAT	CGC	TCC	TTC	TGC	AAG	769
Glu	Leu	Ala	Glu	His	Leu	Thr	Tyr	Leu	Glu	Tyr	Arg	Ser	Phe	Cys	Lys	
			160					165					170			
ATC	CTG	TTT	CAG	GAC	TAT	CAC	AGT	TTC	GTG	ACT	CAT	GGC	TGC	ACT	GTG	817
Ile	Leu	Phe	Gln	Asp	Tyr	His	Ser	Phe	Val	Thr	His	Gly	Cys	Thr	Val	
		175					180					185				
GAC	AAC	CCC	GTC	CTG	GAG	CGG	TTC	ATC	TCC	CTC	TTC	AAC	AGC	GTC	TCA	865
Asp	Asn	Pro	Val	Leu	Glu	Arg	Phe	Ile	Ser	Leu	Phe	Asn	Ser	Val	Ser	
	190					195					200					
CAG	TGG	GTG	CAG	CTC	ATG	ATC	CTC	AGC	AAA	CCC	ACA	GCC	CCG	CAG	CGG	913
Gln	Trp	Val	Gln	Leu	Met	Ile	Leu	Ser	Lys	Pro	Thr	Ala	Pro	Gln	Arg	
205				210						215					220	
GCC	CTG	GTC	ATC	ACA	CAC	TTT	GTC	CAC	GTG	GCG	GAG	AAG	CTG	CTA	CAG	961
Ala	Leu	Val	Ile	Thr	His	Phe	Val	His	Val	Ala	Glu	Lys	Leu	Leu	Gln	
			225						230					235		
CTG	CAG	AAC	TTC	AAC	ACG	CTG	ATG	GCA	GTG	GTC	GGG	GGC	CTG	AGC	CAC	1009
Leu	Gln	Asn	Phe	Asn	Thr	Leu	Met	Ala	Val	Val	Gly	Gly	Leu	Ser	His	
			240					245					250			
AGC	TCC	ATC	TCC	CGC	CTC	AAG	GAG	ACC	CAC	AGC	CAC	GTT	AGC	CCT	GAG	1057
Ser	Ser	Ile	Ser	Arg	Leu	Lys	Glu	Thr	His	Ser	His	Val	Ser	Pro	Glu	
		255					260					265				
ACC	ATC	AAG	CTC	TGG	GAG	GGT	CTC	ACG	GAA	CTA	GTG	ACG	GCG	ACA	GGC	1105
Thr	Ile	Lys	Leu	Trp	Glu	Gly	Leu	Thr	Glu	Leu	Val	Thr	Ala	Thr	Gly	
	270				275						280					
AAC	TAT	GGC	AAC	TAC	CGG	CGT	CGG	CTG	GCA	GCC	TGT	GTG	GGC	TTC	CGC	1153
Asn	Tyr	Gly	Asn	Tyr	Arg	Arg	Arg	Leu	Ala	Ala	Cys	Val	Gly	Phe	Arg	
285				290					295						300	
TTC	CCG	ATC	CTG	GGT	GTG	CAC	CTC	AAG	GAC	CTG	GTG	GCC	CTG	CAG	CTG	1201
Phe	Pro	Ile	Leu	Gly	Val	His	Leu	Lys	Asp	Leu	Val	Ala	Leu	Gln	Leu	
			305						310					315		
GCA	CTG	CCT	GAC	TGG	CTG	GAC	CCA	GCC	CGG	ACC	CGG	CTC	AAC	GGG	GCC	1249
Ala	Leu	Pro	Asp	Trp	Leu	Asp	Pro	Ala	Arg	Thr	Arg	Leu	Asn	Gly	Ala	
			320				325						330			
AAG	ATG	AAG	CAG	CTC	TTT	AGC	ATC	CTG	GAG	GAG	CTG	GCC	ATG	GTG	ACC	1297
Lys	Met	Lys	Gln	Leu	Phe	Ser	Ile	Leu	Glu	Glu	Leu	Ala	Met	Val	Thr	
		335					340					345				



AGC Ser	CTG Leu	CGG Arg	CCA Pro	CCA Pro	GTA Val	CAG Gln	AAC Asn	CCC Pro	GAC Asp	CTG Leu	CTG Leu	AGC Ser	CTC Leu	CTC Leu	1345	
350						355				360						
ACG Thr	GTG Val	TCT Ser	CTG Leu	GAT Asp	CAG Gln	TAT Tyr	CAG Gln	ACG Thr	GAG Glu	GAT Asp	GAG Glu	CTG Leu	TAC Tyr	CAG Gln	CTG Leu	1393
365					370					375					380	
TCC Ser	CTG Leu	CAG Gln	CGG Arg	GAG Glu	CCG Pro	CGC Arg	TCC Ser	AAG Lys	TCC Ser	TCG Ser	CCA Pro	ACC Thr	AGC Ser	CCC Pro	ACG Thr	1441
				385					390					395		
AGT Ser	TGC Cys	ACC Thr	CCA Pro	CCA Pro	CCC Pro	CGG Arg	CCC Pro	CCG Pro	GTA Val	CTG Leu	GAG Glu	GAG Glu	TGG Trp	ACC Thr	TCG Ser	1489
			400					405					410			
GCT Ala	GCC Ala	AAA Lys	CCC Pro	AAG Lys	CTG Leu	GAT Asp	CAG Gln	GCC Ala	CTC Leu	GTG Val	GTG Val	GAG Glu	CAC His	ATC Ile	GAG Glu	1537
		415					420					425				
AAG Lys	ATG Met	GTG Val	GAG Glu	TCT Ser	GTG Val	TTC Phe	CGG Arg	AAC Asn	TTT Phe	GAC Asp	GTC Val	GAT Asp	GGG Gly	GAT Asp	GGC Gly	1585
		430				435					440					
CAC His	ATC Ile	TCA Ser	CAG Gln	GAA Glu	GAA Glu	TTC Phe	CAG Gln	ATC Ile	ATC Ile	CGT Arg	GGG Gly	AAC Asn	TTC Phe	CCT Pro	TAC Tyr	1633
445					450					455					460	
CTC Leu	AGC Ser	GCC Ala	TTT Phe	GGG Gly	GAC Asp	CTC Leu	GAC Asp	CAG Gln	AAC Asn	CAG Gln	GAT Asp	GGC Gly	TGC Cys	ATC Ile	AGC Ser	1681
				465					470					475		
AGG Arg	GAG Glu	GAG Glu	ATG Met	GTT Val	TCC Ser	TAT Tyr	TTC Phe	CTG Leu	CGC Arg	TCC Ser	AGC Ser	TCT Ser	GTG Val	TTG Leu	GGG Gly	1729
			480					485					490			
GGG Gly	CGC Arg	ATG Met	GGC Gly	TTC Phe	GTA Val	CAC His	AAC Asn	TTC Phe	CAG Gln	GAG Glu	AGC Ser	AAC Asn	TCC Ser	TTG Leu	CGC Arg	1777
		495					500					505				
CCC Pro	GTC Val	GCC Ala	TGC Cys	CGC Arg	CAC His	TGC Cys	AAA Lys	GCC Ala	CTG Leu	ATC Ile	CTG Leu	GGC Gly	ATC Ile	TAC Tyr	AAG Lys	1825
	510					515					520					
CAG Gln	GGC Gly	CTC Leu	AAA Lys	TGC Cys	CGA Arg	GCC Ala	TGT Cys	GGA Gly	GTG Val	AAC Asn	TGC Cys	CAC His	AAG Lys	CAG Gln	TGC Cys	1873
525					530					535					540	
AAG Lys	GAT Asp	CGC Arg	CTG Leu	TCA Ser	GTT Val	GAG Glu	TGT Cys	CGG Arg	CGC Arg	AGG Arg	GCC Ala	CAG Gln	AGT Ser	GTG Val	AGC Ser	1921
				545					550					555		
CTG Leu	GAG Glu	GGG Gly	TCT Ser	GCA Ala	CCC Pro	TCA Ser	CCC Pro	TCA Ser	CCC Pro	ATG Met	CAC His	AGC Ser	CAC His	CAT His	CAC His	1969
			560					565					570			
CGC Arg	GCC Ala	TTC Phe	AGC Ser	TTC Phe	TCT Ser	CTG Leu	CCC Pro	CGC Arg	CCT Pro	GGC Gly	AGG Arg	CGA Arg	GGC Gly	TCC Ser	AGG Arg	2017
		575					580					585				
CCT Pro	CCA Pro	GAG Glu	ATC Ile	CGT Arg	GAG Glu	GAG Glu	GTA Val	CAG Gln	ACG Thr	GTG Val	GAG Glu	GAT Asp	GGG Gly	GTG Val		2065

- 70 -

TGGAGAAAAT ACTTCAACCA GAGCAGGGAG CCTGGGGGTG TCGGGGCAGG AGGCTGGGGA 2180  
 TGGGGGTGGG ATATGAGGGT GGCATGCAGC TGAGGGCAGG GCCAGGGCTG GTGTCCCTAA 2240  
 GGTGTACAG ACTCTTGTGA ATATTTGTAT TTTCCAGATG GAATAAAAAG GCCCGTGTA 2300  
 TTAACCTTC 2309

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 609 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
 1 5 10 15  
 Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly Lys Val Arg Asp  
 20 25 30  
 Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro Trp Tyr Ile Pro  
 35 40 45  
 Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr Gln Gln Ser Arg  
 50 55 60  
 Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys His Leu Val Arg  
 65 70 75 80  
 Tyr Trp Ile Ser Ala Phe Pro Ala Glu Phe Asp Leu Asn Pro Glu Leu  
 85 90 95  
 Ala Glu Gln Ile Lys Glu Leu Lys Ala Leu Leu Asp Gln Glu Gly Asn  
 100 105 110  
 Arg Arg His Ser Ser Leu Ile Asp Ile Asp Ser Val Pro Thr Tyr Lys  
 115 120 125  
 Trp Lys Arg Gln Val Thr Gln Arg Asn Pro Val Gly Gln Lys Lys Arg  
 130 135 140  
 Lys Met Ser Leu Leu Phe Asp His Leu Glu Pro Met Glu Leu Ala Glu  
 145 150 155 160  
 His Leu Thr Tyr Leu Glu Tyr Arg Ser Phe Cys Lys Ile Leu Phe Gln  
 165 170 175  
 Asp Tyr His Ser Phe Val Thr His Gly Cys Thr Val Asp Asn Pro Val  
 180 185 190  
 Leu Glu Arg Phe Ile Ser Leu Phe Asn Ser Val Ser Gln Trp Val Gln  
 195 200 205  
 Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg Ala Leu Val Ile  
 210 215 220  
 Thr His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe  
 225 230 235 240  
 Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser  
 245 250 255

- 71 -

Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu Thr Ile Lys Leu  
 260 265 270  
 Trp Glu Gly Leu Thr Glu Leu Val Thr Ala Thr Gly Asn Tyr Gly Asn  
 275 280 285  
 Tyr Arg Arg Arg Leu Ala Ala Cys Val Gly Phe Arg Phe Pro Ile Leu  
 290 295 300  
 Gly Val His Leu Lys Asp Leu Val Ala Leu Gln Leu Ala Leu Pro Asp  
 305 310 315 320  
 Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala Lys Met Lys Gln  
 325 330 335  
 Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr Ser Leu Arg Pro  
 340 345 350  
 Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu Thr Val Ser Leu  
 355 360 365  
 Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu Ser Leu Gln Arg  
 370 375 380  
 Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr Ser Cys Thr Pro  
 385 390 395 400  
 Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser Ala Ala Lys Pro  
 405 410 415  
 Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu Lys Met Val Glu  
 420 425 430  
 Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly His Ile Ser Gln  
 435 440 445  
 Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr Leu Ser Ala Phe  
 450 455 460  
 Gly Asp Leu Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met  
 465 470 475 480  
 Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly Gly Arg Met Gly  
 485 490 495  
 Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys  
 500 505 510  
 Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys  
 515 520 525  
 Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu  
 530 535 540  
 Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser Leu Glu Gly Ser  
 545 550 555 560  
 Ala Pro Ser Pro Ser Pro Met His Ser His His His Arg Ala Phe Ser  
 565 570 575  
 Phe Ser Leu Pro Arg Pro Gly Arg Arg Gly Ser Arg Pro Pro Glu Ile  
 580 585 590  
 Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val Phe Asp Ile His  
 595 600 605  
 Leu

- 72 -

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 832 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 11..733

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCCCGCCGCC	ATG	CCG	CCC	TTA	CTG	CCC	CTG	CGC	CTG	TGC	CGG	CTG	TGG	49
	Met	Pro	Pro	Leu	Leu	Pro	Leu	Arg	Leu	Cys	Arg	Leu	Trp	
	1				5					10				
CCC CGC AAC CCT CCC TCC CGG CTC CTC GGA GCG GCC GCC GGG CAG CGG														97
Pro Arg Asn Pro Pro Ser Arg Leu Leu Gly Ala Ala Ala Gly Gln Arg														
	15				20				25					
TCC AGA CCC AGT ACT TAT TAT GAA CTG TTG GGG GTG CAT CCT GGT GCC														145
Ser Arg Pro Ser Thr Tyr Tyr Glu Leu Leu Gly Val His Pro Gly Ala														
	30				35			40					45	
AGC ACT GAG GAA GTT AAA CGA GCT TTC TTC TCC AAG TCC AAA GAG CTG														193
Ser Thr Glu Glu Val Lys Arg Ala Phe Phe Ser Lys Ser Lys Glu Leu														
				50				55					60	
CAC CCA GAC CGG GAC CCT GGG AAC CCA AGC CTG CAC AGC CGC TTT GTG														241
His Pro Asp Arg Asp Pro Gly Asn Pro Ser Leu His Ser Arg Phe Val														
				65			70					75		
GAG CTG AGC GAG GCA TAC CGT GTG CTC AGC CGT GAG CAG AGC CGC CGC														289
Glu Leu Ser Glu Ala Tyr Arg Val Leu Ser Arg Glu Gln Ser Arg Arg														
			80				85					90		
AGC TAT GAT GAC CAG CTC CGC TCA GGT AGT CCC CCA AAG TCT CCA CGA														337
Ser Tyr Asp Asp Gln Leu Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg														
			95			100				105				
ACC ACA GTC CAT GAC AAG TCT GCC CAC CAA ACA CAC AGC TCC TGG ACA														385
Thr Thr Val His Asp Lys Ser Ala His Gln Thr His Ser Ser Trp Thr														
	110				115				120				125	
CCC CCC AAC GCA CAG TAC TGG TCC CAG TTT CAC AGC GTG AGG CCA CAG														433
Pro Pro Asn Ala Gln Tyr Trp Ser Gln Phe His Ser Val Arg Pro Gln														
			130					135					140	
GGG CCC CAG TTG AGG CAG CAG CAA CAC AAA CAA AAC AAA CAA GTG CTG														481
Gly Pro Gln Leu Arg Gln Gln Gln His Lys Gln Asn Lys Gln Val Leu														
			145				150					155		
GGG TAC TGC CTC CTC CTC ATG CTG GCG GGC ATG GGC CTG CAC TAC ATT														529
Gly Tyr Cys Leu Leu Leu Met Leu Ala Gly Met Gly Leu His Tyr Ile														
			160			165				170				
GCC TTC AGG AAG GTG AAG CAG ATG CAC CTT AAC TTC ATG GAT GAA AAG														577
Ala Phe Arg Lys Val Lys Gln Met His Leu Asn Phe Met Asp Glu Lys														
	175				180				185					

- 73 -

GAT CGG ATC ATC ACA GCC TTC TAC AAC GAA GCC CGG GCA CGG GCC AGG 625  
 Asp Arg Ile Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg  
 190 195 200 205  
 GCC AAC AGA GGC ATC CTT CAG CAG GAG CGA CAA CGG CTA GGG CAG CGG 673  
 Ala Asn Arg Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg  
 210 215 220  
 CAG CCG CCA CCA TCC GAG CCA ACC CAA GGC CCC GAG ATC GTG CCC CGG 721  
 Gln Pro Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg  
 225 230 235  
 GGC GCC GGC CCC TGA GGGGCTC ACCTGGATGG GGCCTGCAGT GCGTTCCCGC 773  
 Gly Ala Gly Pro \*  
 240  
 TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTCGCAG 832

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 241 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Pro Pro Leu Leu Pro Leu Arg Leu Cys Arg Leu Trp Pro Arg Asn  
 1 5 10 15  
 Pro Pro Ser Arg Leu Leu Gly Ala Ala Gly Gln Arg Ser Arg Pro  
 20 25 30  
 Ser Thr Tyr Tyr Glu Leu Leu Gly Val His Pro Gly Ala Ser Thr Glu  
 35 40 45  
 Glu Val Lys Arg Ala Phe Phe Ser Lys Ser Lys Glu Leu His Pro Asp  
 50 55 60  
 Arg Asp Pro Gly Asn Pro Ser Leu His Ser Arg Phe Val Glu Leu Ser  
 65 70 75 80  
 Glu Ala Tyr Arg Val Leu Ser Arg Glu Gln Ser Arg Arg Ser Tyr Asp  
 85 90 95  
 Asp Gln Leu Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg Thr Thr Val  
 100 105 110  
 His Asp Lys Ser Ala His Gln Thr His Ser Ser Trp Thr Pro Pro Asn  
 115 120 125  
 Ala Gln Tyr Trp Ser Gln Phe His Ser Val Arg Pro Gln Gly Pro Gln  
 130 135 140  
 Leu Arg Gln Gln Gln His Lys Gln Asn Lys Gln Val Leu Gly Tyr Cys  
 145 150 155 160  
 Leu Leu Leu Met Leu Ala Gly Met Gly Leu His Tyr Ile Ala Phe Arg  
 165 170 175  
 Lys Val Lys Gln Met His Leu Asn Phe Met Asp Glu Lys Asp Arg Ile  
 180 185 190  
 Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg Ala Asn Arg

- 74 -

195	200	205
Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg Gln Pro Pro		
210	215	220
Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg Gly Ala Gly		
225	230	235 240
Pro		

SEQ ID Nos: 10-18 25-36

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 300 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 170..300

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGATTTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG	60
CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC	120
TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAG CCC CAT	175
	Pro His
	1
GGG AAC GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC	223
Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser	
5 10 15	
CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC	271
Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp	
20 25 30	
CTG GAC AAG GGC TGC ACG GTG GAG GAG CT	300
Leu Asp Lys Gly Cys Thr Val Glu Glu Leu	
35 40	

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 75 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Pro His Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu  
 1 5 10 15  
 Arg Ser Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr  
 20 25 30  
 Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
 35 40

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGATCCCCC TGGTC

15

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asp Val Asp Glu Glu Asp Glu Val Glu Asp Ile Glu Phe  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids

- 76 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asp His Asp Arg Asp Gly Phe Ile Ser Gln Glu Glu Phe  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Val Asp Met Asp Gly Gln Ile Ser Lys Asp Glu Leu  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe Asn  
 1 5 10 15  
 Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser Arg  
 20 25 30



- 77 -

Leu Lys Glu Thr His  
35

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Phe Val His Val Ala Lys His Leu Arg Lys Ile Asn Asn Phe Asn  
1 5 10 15  
 Thr Leu Met Ser Val Val Gly Gly Ile Thr His Ser Ser Val Ala Arg  
20 25 30  
 Leu Ala Lys Thr Tyr  
35

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys Arg His  
1 5 10 15  
 Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg  
20 25 30  
 Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val  
35 40 45  
 Glu Cys  
50

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

His Asn Phe His Glu Thr Thr Phe Leu Thr Pro Thr Thr Cys Asn His

- 78 -

1	5	10	15
Cys	Asn	Lys	Leu
	20	Leu	Trp
		Gly	Ile
		25	Leu
		Arg	Gln
		Gly	Phe
			30
		Lys	Cys
		Lys	
Asp	Cys	Gly	Leu
	35	Ala	Val
		His	Ser
		40	Cys
		Cys	Lys
		Ser	Asn
			45
		Ala	Val
		Ala	
Glu	Cys		
	50		

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGGATCCCCC TGGTC

15

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAATTCGGCA CGAGCCGACG G

21

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 78 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGGAGCAGA AGCTGATCTC CGAGGAGGAC CTGCCCCGGG CAGCTGGATC CGCAGCCCAC

60

CCCGCGCCCG CGGCCATG

78

## (2) INFORMATION FOR SEQ ID NO:22:

- 79 -

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 26 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Glu Gln Lys Leu Ile. Ser Glu Glu Asp Leu Pro Gly Ala Ala Gly  
1                    5                    10                    15  
Ser Ala Ala His Pro Ala Pro Ala Ala Met  
                    20                    25

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 33 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGATCCGCAG CCCACCCCGC GCCGGCGGCC ATG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 11 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met  
                    5                    10

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 21 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 80 -

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:  
GGACAAAGTG TGTGATGAAC C 21
- (2) INFORMATION FOR SEQ ID NO:26:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:  
CTCATCCTCC GTCTGATACT G 21
- (2) INFORMATION FOR SEQ ID NO:27:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:  
GTAGATGTGG ATCAGCTTGG 20
- (2) INFORMATION FOR SEQ ID NO:28:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:  
AGGTGGAGAA TGGTCAAGG 19
- (2) INFORMATION FOR SEQ ID NO:29:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:  
GTCATAGTCT GTCTCCTACT 20

- 81 -

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ACATAGACAG CGTGCCTACC

20

## (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TACAACCTTA GGGACACCAG

20

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TGCTGAGCCT GCTCACGGTG

20

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CAAGTGAACA GCACGTCC

18

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:

- 82 -

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GACTATCTCA AGGACCAGCT G

21

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGTTCGGTCC GAGCCCGG

18

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGAGCGATAC TCCAAGTAGG T

21

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AGCGGGCCAG GCCCCTTC

18

(2) INFORMATION FOR SEQ ID NO:38:

- 83 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATCCTGGTC CAATGCGCTC

20

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCACTGAGGA AGTTAAACGA GC

22

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCTCGTTTAA CTTCTCAGT GC

22

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCTCAGCTCC ACAAAGCGGC T

21

(2) INFORMATION FOR SEQ ID NO:42:

- 84 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ACCAGCTCCG CTCAGGTAG

19

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TCCAGGAGCT GTGTGTTTGG

20

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CCAGTTTCAC AGCGTGAGG

19

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CAGCATGAGG AGGAGGCAG

19



## CLAIMS:

1. An isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
2. An isolated nucleic acid molecule according to claim 1 wherein the regulator comprises a zinc finger domain of an  $(\text{HC}_3)_2$  type.
3. An isolated nucleic acid molecule according to claim 2 wherein the sequence of nucleotides or complementary sequence of nucleotides is selected from:
  - (i) a nucleotide sequence set forth in SEQ ID NO:2;
  - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
  - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
  - (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
4. An isolated nucleic acid molecule according to claim 1 wherein said gene regulator is a guanine nucleotide exchange factor (GEF) or a derivative thereof.
5. An isolated nucleic acid molecule according to claim 4 wherein the sequence of nucleotides is selected from:
  - (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
  - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
  - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
  - (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the

- 86 -

nucleotide sequence set forth in (i), (ii) or (iii).

6. An isolated nucleic acid molecule according to claim 1, wherein said gene regulator is a heat shock protein or is a heat shock binding protein or a derivative thereof.

7. An isolated nucleic acid molecule according to claim 6, wherein the sequence of nucleotides is selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

8. A genetic construct comprising a vector portion and a gene portion comprising a regulator of gene expression or a derivative thereof .

9. A genetic construct according to claim 8 wherein the gene portion comprises a zinc finger domain of  $(\text{HC}_3)_2$  type.

10. A genetic construct according to claim 9 wherein the gene portion comprises a nucleotide sequence selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

- 87 -

11. A genetic construct according to claim 8 wherein said gene portion is a nucleotide exchange factor (GEF) or derivative thereof.
12. A genetic construct according to claim 11 wherein the gene portion comprises a nucleotide sequence selected from:
  - (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
  - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
  - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
  - (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
13. A genetic construct according to claim 8 wherein the gene portion is a heat shock protein or a derivative thereof or a heat shock binding protein or derivative thereof.
14. A genetic construct according to claim 13 wherein the gene portion comprises a nucleotide sequence selected from:
  - (i) a nucleotide sequence set forth in SEQ ID NO:8;
  - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
  - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
  - (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
15. A nucleic acid molecule encoding a gene regulator having the identifying characteristics of a molecule selected from MCG4, MCG7 and MCG18 having respective amino acid sequences of SEQ ID NO:3, SEQ ID NO: 5 or 7 and SEQ ID NO:9.

16. A method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg4* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

17. A method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

18. A method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

19. A method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg7* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

20. A method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

21. A method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

22. A method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

23. A method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

24. A method for detecting MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

1/85

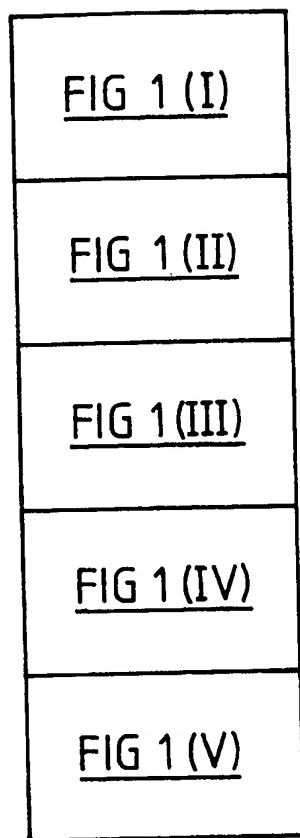


FIG 1

2/85

**FIGURE 1 (I)**

[illegible]

3/85

**FIGURE 1 (II)**

GCT	GCC	CAG	CTA	CCC	CGA	AAC	ACG	GCA	CCT	GCC	GGC	TAT	CAG	TGC	320
Ala	Ala	Gln	Leu	Pro	Arg	Asn	Thr	Ala	Pro	Ala	Gly	Tyr	Gln	Cys	
							90					95			
CCC	AGC	TGC	AAT	GGC	CCC	ATC	TTC	CCC	CCA	ACC	AAC	CTG	GCT	GGC	365
Pro	Ser	Cys	Asn	Gly	Pro	Ile	Phe	Pro	Pro	Thr	Asn	Leu	Ala	Gly	
							105					110			
CCC	GTG	GCC	TCC	GCA	CTG	AGA	GAG	AAG	CTG	GCC	ACA	GTC	AAC	TGG	410
Pro	Val	Ala	Ser	Ala	Leu	Arg	Glu	Lys	Leu	Ala	Thr	Val	Asn	Trp	
							120					125			
GCC	CGG	GCA	GGA	CTG	GGC	CTC	CCT	CTG	ATC	GAT	GAG	GTG	GTG	AGC	455
Ala	Arg	Ala	Gly	Leu	Gly	Leu	Pro	Leu	Ile	Asp	Glu	Val	Val	Ser	
							135					140			
CCA	GAG	CCC	GAG	CCC	CTC	AAC	ACG	TCT	GAC	TTC	TCT	GAC	TGG	TCT	500
Pro	Glu	Pro	Glu	Pro	Leu	Asn	Thr	Ser	Asp	Phe	Ser	Asp	Trp	Ser	
							150					155			
AGT	TTT	AAT	GCC	AGC	AGT	ACC	CCT	GGA	CCA	GAG	GAG	GTA	GAC	AGC	545
Ser	Phe	Asn	Ala	Ser	Ser	Thr	Pro	Gly	Pro	Glu	Glu	Val	Asp	Ser	
							165					170			



4/85

**FIGURE 1 (III)**

GCC TCT	GCT GCC	CCA GCC	TTC	TAC	AGC	CGA	GCC	CCC	CGG	CCC	CCA	590
Ala Ser	Ala Ala	Pro Ala	Phe	Tyr	Ser	Arg	Ala	Pro	Arg	Pro	Pro	
175	180								185			
GCT TCC	CCA GGC	CGG CCC	GAG	CAG	CAC	ACA	GTG	ATC	CAC	ATG	GGC	635
Ala Ser	Pro Gly	Arg Pro	Glu	Gln	His	Thr	Val	Ile	His	Met	Gly	
190	195								200			
AAT CCT	GAG CCC	TTG ACT	CAC	GCC	CCT	AGG	AAG	GTG	TAT	GAT	ACG	680
Asn Pro	Glu Pro	Leu Thr	His	Ala	Pro	Arg	Lys	Val	Tyr	Asp	Thr	
205	210								215			
CGG GAT	GAT GAC	CGG ACA	CCA	GGC	CTC	CAT	GGA	GAC	TGT	GAC	GAT	725
Arg Asp	Asp Asp	Arg Arg	Thr Pro	Gly Pro	Leu	His	Gly	Asp	Cys	Asp	Asp	
220	225								230			
GAC AAG	TAC CGA	CGT CGG	CCG	GCC	TTG	GGT	TGG	CTG	GCC	CGG	CTG	770
Asp Lys	Tyr Arg	Arg Arg	Pro Pro	Ala Ala	Leu	Gly	Trp	Leu	Ala	Arg	Leu	
235	240								245			
CTA AGG	AGC CGG	GCT GGG	TCT	CGG	AAG	CGG	CCG	CTG	ACC	CTG	CTC	815
Leu Arg	Ser Arg	Ala Gly	Ser	Arg	Lys	Arg	Pro	Leu	Thr	Leu	Leu	
250	255								260			

5/85

**FIGURE 1 (IV)**

CAG CGG GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG	860
Gln Arg Ala Gly Leu Leu Leu Leu Leu Gly Leu Gly Phe Leu	
265	270
GCC CTC CTT GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC	905
Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Asp	
280	285
AGC GAT CCC AAC CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG	950
Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His Ile Arg Val	
295	300
GGC CCC TCC TGA GCGCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT	1002
Gly Pro Ser *	
310	
CTGTGGAGGA GAGCGGGGT AATGGGGAGG CTGAGGGCAC CTCCTCACTG	1052
CCCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT	1102
CCACCAGAGT GGCTCGCAGG CCAGGCCCTGG AGTCCCCCGTG GGTC AAGCAT	1152

6/85

**FIGURE 1 (V)**

TTGTCTTGAC TTGCTTTCTC CCGGTCTCC AGCCTCCGAC CCCTCGCCCC	1202
ATGAAGGAGC TGGCAGGTGG AAATAAACAA CAACTTTATT	1242

**FIGURE 2**

gb|AA155210|AA155210 mr98e01.r1 Stratagene mouse embryonic carcinoma  
(#937317)Mus musculus cDNA clone 605496 5'

Query: 1 MGLCKCPKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNIP 60  
MGLCKCPKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCN PL  
Sbjct: 98 MGLCKCPKRKVTNLFCEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNTPL 277

7/85

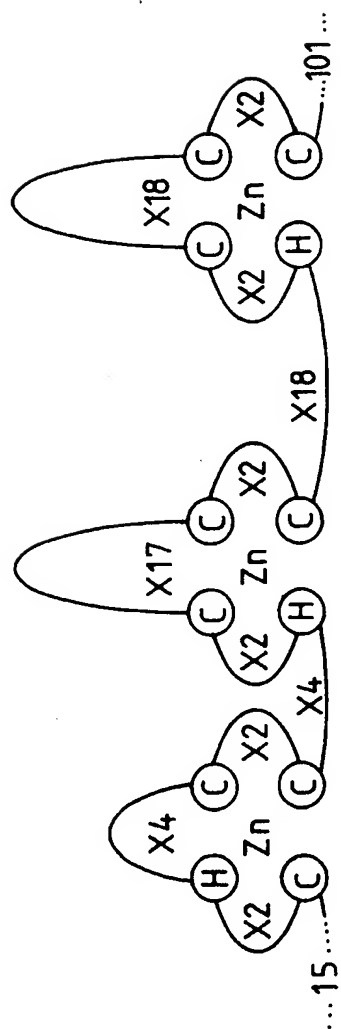
**FIGURE 3**

dbj|D75913| CELK111G3F C. elegans cDNA clone ykl11lg3:5' end, single read.

8/85

Query:	7	PKRKVTNLF	CFEHRVNV	CEHCLVAN	HAKCIVQSY	LQWLQD	SDYNPN	CRCLNI	PLASRETT	66
		PKRKVTNLF	+EHRVNV	CE LV NH	C+VQSYL	WL D	DY+PNC	LC L	+T	
Sbjct:	1	PKRKVTNLF	XYEHRVNV	CELXLVDN	HPNCVVQSY	LTWLT	DDQDYD	PNC	SLCKTTLXEGDTI	180
Query:	67	RLVCYDL	FHWACLN	ERAAQL	PRNTAP	AGYQCP	98	98	PSCNGPI	FPPNQ 109
		RL C L	HW C +E	P	TAP	GY+CP			P C+ +FPP+Q	
Sbjct:	181	RLNCLHLL	HWKCFDE	WXGNF	PDTTAP	XXGYRCP	276	275	PCCSQEV	FPPDQ 310

9/85

FIG 4

10/85

**FIGURE 5**

sp | P46580 | YLBS\_CAEEL HYPOTHETICAL 146.8 KD PROTEIN C34E10.5 IN  
CHROMOSOME III gi|500728 (U10402) C34E10.5 gene product  
[Caenorhabditis elegans]

Query: 56 CNIPLASRETRRLVCYDLFWACLNERRAAQLPRNTAPAGYQCPSC 100  
C+I L ++ + L C LF W C+ E A + + + +CP C  
Sbjct:1222 CSICLENKNPSALFCGHLFCWTCIQEHAVAATSSASTSSARCPQC 1266

11/85

FIGURE 6

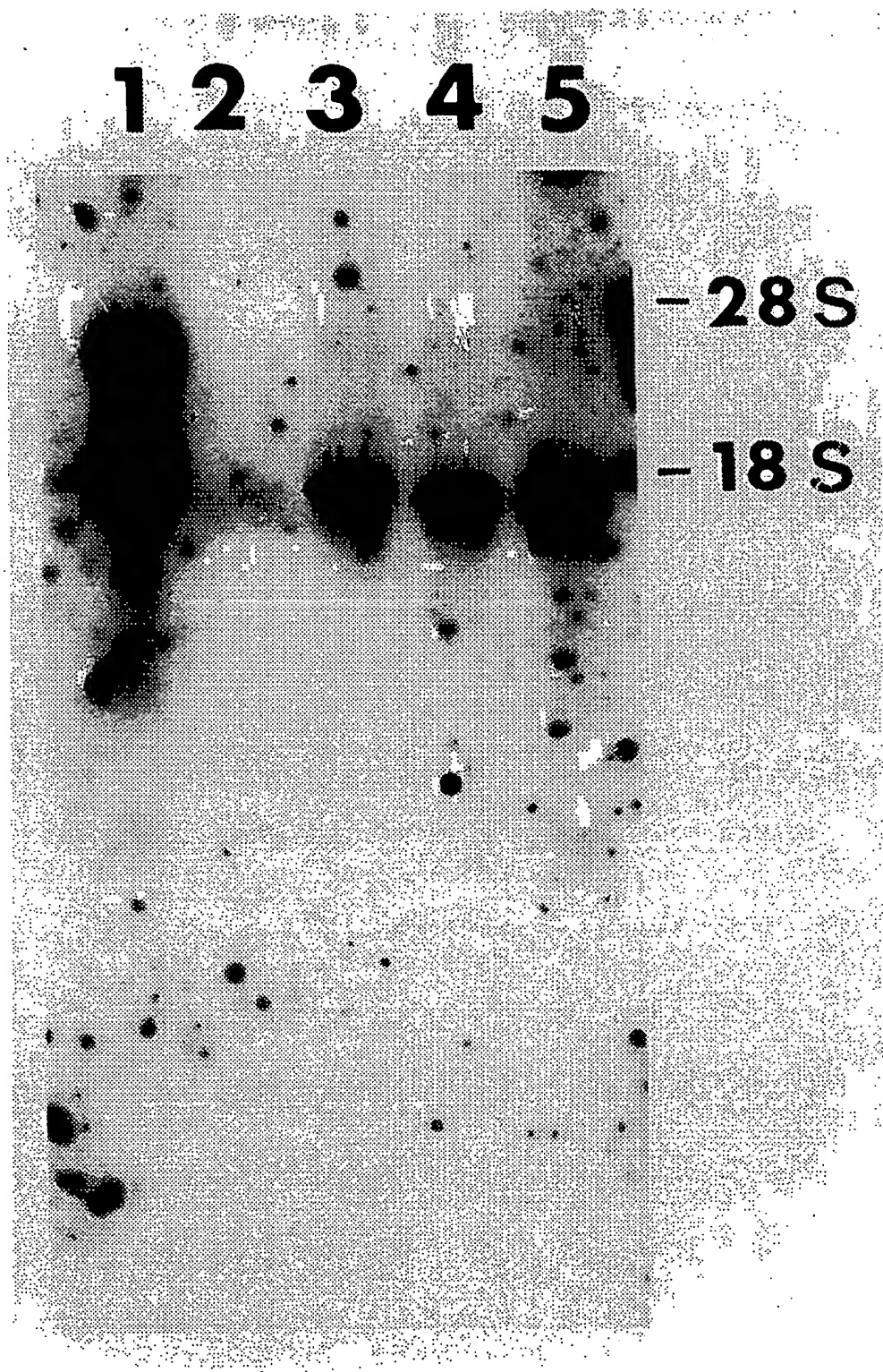
gi | 703468 | (L29051) homologous to GATA-binding transcription factor  
[Schizosaccharomyces pombe]

Query:	35	CIVQSYLQWLQDSDYNPNCRLCNI	58
		C + +W +D NP C C +	
Sbjct:	175	CATTNTPKWRRDESGNPICNACGL	198
Query:	162	SSTPGPEEVDASAAAFYSQAPRPPASPGRPEQHTVTHMGNPEPLTHAPRKVYDTRDDD	221
		+S PEE S S S P+ SP+ +Q +I P +V + D	
Sbjct:	441	ASLLNPEEPPSNSDKQPSMSNGPKSEVSPSQSQQAPLIQSSTSPVSLQFPPEVQGSNVDK	500
Query:	222	RTPGLH	227
		R L+	
Sbjct:	501	RNYALN	506



12/85

FIG 7



13/85

<u>FIG 8 (I)</u>
<u>FIG 8 (II)</u>
<u>FIG 8 (III)</u>
<u>FIG 8 (IV)</u>
<u>FIG 8 (V)</u>
<u>FIG 8 (VI)</u>
<u>FIG 8 (VII)</u>

FIG 8

**FIGURE 8 (I)**

gb | AA074703|AA074703 zm76g07.r1 Stratagene neuroepithelium (#937231)

Homo sapiens cDNA clone 531612 5'

Length = 417

Plus Strand HSPs:

Score = 818 (226.0 bits), Expect = 6.1e-103, Sum P(5)=6.1e-103

Identities = 206/259 (79%), Positives = 206/259 (79%), Strand = Plus/Plus

14/85

Query: 446 GGCTCCCTCTGATCGATGAGGTGGTGAGCCAGAGCCCGAGCCCCCTCAACACGTCAGAC 505

|||||

Sbjct: 49 GGGCTCCCTCTGATCGATGAGGTGATAAGCCAGAGCCCGAGCCCCCTCAATTCCCTCAGAC 108

Query: 506 TTCCTGACTGGTCTAGTTTAAATGCCAGCAGTACCCCTGGACCAGAGGAGGTAGACAGC 565

|||||

Sbjct: 109 TTCCTGATTGGTCCAGCTTTAAATGCCACCACCCTCTGTGTGCAAGAGGAGAGAGCCAGC 168

**FIGURE 8 (II)**

15/85

```
Query: 566 GCCTCTGCTGCCCCAGCCTTCTACAGCCAGGCCCGCCCGCCAGCTTCCCCAGGCCGG 625
      |||
Sbjct: 169 ACTCCATCTGCACCTGCTTCTATAGCCAGGCTCCCCGCCCTCCTCCCTCCCTCCCAAGCCGT 228

Query: 626 CCCGAGCAGCACACAGTGATCCACATGGGCAATCCTGAGCCCTTGACTCAGCCCCCTAGG 685
      |||
Sbjct: 229 CCCGAGCAGCACACAGTCAATACACATGGGGAGTACTGAAGCCCTGGCACACGCCCCAAGG 288

Query: 686 AAGGTGTATGATACGCCGG 704
      |||
Sbjct: 289 AAAGTATATGACACACCCGG 307
```

Score = 230 (63.6 bits), Expect = 6.1e-103, Sum P(5)=6.1e-103  
Identities = 50/55 (90%), Positives = 50/55 (90%), Strand = Plus/Plus

**FIGURE 8 (III)**

Query: 398 GCACTGAGAGAGAAGCTGGCCACAGTCAACTGGGCCCGGCAGGACTGGGCTCC 452  
|||||  
Sbjct: 2 GCACTGAGAGAAAAGCTAGCCACAGTCAACTGGGCCCGGCAGGACTGGGCTCCC 56

16/85

Score = 175 (48.4 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103  
Identities = 39/44 (88%), Positives = 39/44 (88%), Strand = Plus/Plus

Query: 767 GCCTTGGGTTGGCTGGCCCGGCTGCTAAGGAGCCGGGCTGGGTC 810  
|||  
Sbjct: 373 GCTCTGGGCTGGCTGGCCCGGCTGCTCAGGAGCCGGGCTGGGTC 416

Score = 139 (38.4 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103  
Identities = 31/35 (88%), Positives = 31/35 (88%), Strand = Plus/Plus

17/85

**FIGURE 8 (IV)**

Query: 731 GGAGACTGTGACGATGACAAGTACCGACGTCGGCC 765  
|||||  
Sbjct: 336 GGAGACTGTGATGATGACAAATACCGCCGCGGCC 370

Score = 133 (36.8 bits), Expect = 6.1e-103, Sum P(5)=5.1e-103  
Identities = 29/32 (90%), Positives = 29/32 (90%), Strand = Plus/Plus

Query: 701 CGGGATGATGACCGGACACCGAGGCTCCATGG 732  
|||||  
Sbjct: 305 CGGGATGATGACCGGACACCGAGGATTCATGG 336







20/85

**FIGURE 8 (VII)**

Query: 398 GCACTGAGAGAGAAGCTGGCCACAGTCAACTGGGCCCCGGGCAGGACTGGGGCTCCCTCTG 457  
 |||||  
 Sbjct: 2 GCACTGAGAGACAAGCTAGCCACAGTCAACTGGGCCCCGGGCAGGACTGGGGCTCCCTCTG 61

Query: 458 ATCGATGAGGTGGTGAGCCCCAGAGCCCCGAGCCCCCTCAA 495  
 |||||  
 Sbjct: 62 ATCGATGAGGTGATAAGCCCCAGAGCCCCGAGCCCCCTCAA 99

Score = 219 (60.5 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87  
 Identities = 51/60 (85%), Positives = 51/60 (85%), Strand = Plus/Plus

Query: 702 GGGATGATGACCCGGACACCAGGCCCTCCATGGAGACTGTGACGATGACAAAGTACCGACGTC 761  
 |||||  
 Sbjct: 309 GGATTGATGACCCGGACAGCAGGCATTTCATGGAGACTGTGATGATGACAAATACCGCCGCC 368

**FIGURE 9**

W32939 human

TACCGCCCTTCGGAACCAAGTGCAGCGGCCGATCAGTAAACACAGAGACTGGGATCGATCATGGGGCTTTGTAAG

AA242159 mouse

CTTCCGGCGCTTTTCATTACCGTACGCACCGGTCA-CGATCGGCATCGCGGAGGATCGGTCATGGGACTTTTGCAAG

21/85

22/85

<u>FIG 10 (I)</u>
<u>FIG 10 (II)</u>
<u>FIG 10 (III)</u>
<u>FIG 10 (IV)</u>

FIG 10

23/85

**FIGURE 10(I)**

MCG4	MGLCKCPKPK	VTNLFCFEHR	VNVCEHCLVA	NHAKCIVQSY	LQWLQDSQDYN	PNCRLCNIP	60
MCG4	ASRETRRLVC	YDLFWACLN	ERAAQLPRNT	APAGYQCPCSC	NGPIFPPTNL	AGPVASALRE	120

3.  $\begin{matrix} \wedge \\ * \\ * \\ * \\ * \end{matrix}$ 

5. [ 74 ] \*\*\*\*\*^

	130	140	150	160	170	180
MCG4						
	*	*	*	*	*	*
	KLATVNWARA	GLGLPLIDEV	VSPEPEPLNT	SDFSDWSSFN	ASSTPGPEEV	DSASAAPAFY
		20	30	40	50	60
1.	*****	i*****s	*****	*tt*svq**r	a*tps*****>	
[ 372 ]	_____					
		30	40	50	60	
2.						
		aq*s*sip	*****	*tt*svq**r	a*tps*****>	
[ 243 ]	_____					

```

3.
[ 229 ]
*****
***** i*****s xrl1*lvql* chhhlcharge sqh*icac*1>
10      20      30      40      50      60      |
|

```

\*

**Abstract**



25/85

FIGURE 10 (III)

5. [ 74 ] arl\*allppq av\*sstqsyw\*vlk\*w-\*t \*qgk\*m\*\*\*\*\* \*\*a\*i\*\*>

g

6. [ 38 ] \_\_\_\_\_ \*t \*q\*\*\*\*\*>

MCG4 LGWLARLLRS RAGSRKRPLT LLQRAGLLLL LGLLGFALL ALMSRLGRAA ADSDPNLDPL

1. [ 372 ] \*\*\*\*\*q\*\*\*\*\* \*\*>

4. [ 86 ] s\*-\*\*>

310

\*

MCG4 MNPHIRVGPS

26/85

**FIGURE 10 (IV)**

Search Analysis for Sequence: MCG4

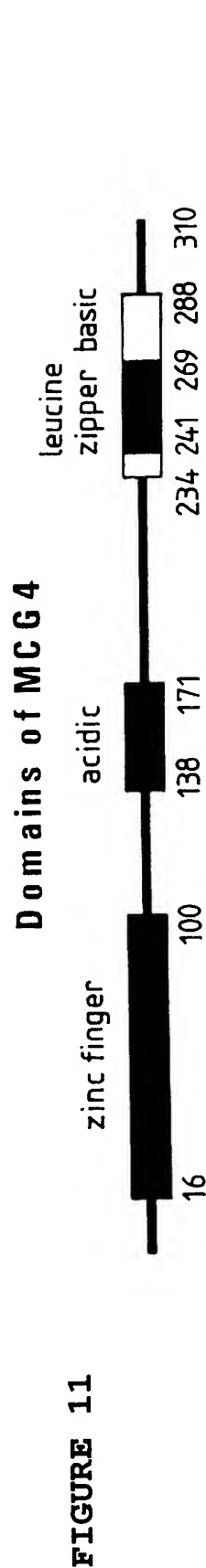
Search from 1 to 310

Date: September 22, 1997

Matrix: pam250 matrix  
 Score Region from 1 to 310  
 Maximum possible score: 1598

## Aligned sequences:

1. = EST AA074703 phase 1 translation
2. = EST AA134788 phase 3 translation
3. = EST AA134788 phase 2 translation
4. = EST AA074703 phase 3 translation
5. = EST AA074703 phase 2 translation
6. = EST AA134788 phase 1 translation



zinc finger consensus: CX<sub>2</sub>HXCX<sub>4</sub>CX<sub>2</sub>HXCX<sub>4</sub>HXCX<sub>17</sub>CX<sub>2</sub>CX<sub>18</sub>HXCX<sub>2</sub>CX<sub>18</sub>CX<sub>2</sub>C

acidic domain consensus: 9/34 negatively charged amino acids, 0/34 positively charged

basic domain consensus: 13/55 positively charged amino acids, 0/55 negatively charged

leucine zipper domain consensus: LX<sub>6</sub>LX<sub>6</sub>RX<sub>6</sub>LX<sub>6</sub>L

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa 261) LX<sub>6</sub>LXLX<sub>6</sub>LXLX<sub>6</sub>L (aa 286)



28/85

<u>FIG 12 (I)</u>	<u>FIG 12 (II)</u>
<u>FIG 12 (III)</u>	<u>FIG 12 (IV)</u>

FIG 12

29/85

**FIG 12 (I)**

Sequences producing High-scoring Segment Pairs:

gn1 PID e236178	(Z70752 F25B3.3 [Caenorhabditis ele...
gi 1293099	(U53884) aimless RasGEF [Dictyosteli...
gi 1655941	(U67326) Ras-GRF2 [Mus musculus]
pir  s30356	CDC25 protein homolog - yeast (Cand...
sp P43069 CC25_CANAL	CELL DIVISION CONTROL PROTEIN 25
sp P28818 GNRP_RAT	GUANINE NUCLEOTIDE RELEASING PROTEIN...
prf  1814463A	guanine nucleotide-releasing factor ...
pir  B46199	nucleotide-exchange-factor homolog c...
gn1  PID e238680	(X97560) hypothetical protein L1309 ...
pir  s22693	CDC25 protein homolog - mouse/gi 50...
sp P14771 SC25_YEAST	SCD25 PROTEIN /gi 457494 (M26647) SD...
sp P26674 STE6_SCHPO	STE6 PROTEIN /pir s28098 ste6 prote...
pir  s28407	CDC25 protein homolog - mouse
sp P27671 GNRP_MOUSE	GUANINE NUCLEOTIDE RELEASING PROTEIN...
gi 386047	(s62035) Ras-specific guanine nucleo...
sp Q02342 CC25_SACKL	CELL DIVISION CONTROL PROTEIN 25 /pi...
pir  s14177	SCD25 protein - yeast (Saccharomyces...
gi 433720	(L26584) CDC25 [Homo sapiens]
gn1 PID e241744	(Z68880) T14G10.2 [Caenorhabditis el...
gi 3484	(X03579) CDC25 protein (aa 1-1588) [...

30/85

High Score	smallest Sum Probability P(N)	N
307	3.0e-124	8
202	7.8e-22	5
152	3.6e-16	4
150	2.2e-15	3
150	2.2e-15	3
166	2.6e-15	3
166	2.6e-15	3
167	1.1e-14	1
158	3.0e-14	3
167	3.7e-14	2
158	4.6e-14	3
160	5.2e-14	2
167	1.2e-13	3
167	1.2e-13	3
153	2.0e-13	2
142	4.5e-13	2
152	5.7e-13	3
153	6.0e-13	3

FIG 12 (II)

31/85

sp  PO4821 CC25_YEAST	CELL DIVISION CONTROL PROTEIN 25 /pi...
gi  915328	(U24070) Munc13-1 [Rattus norvegicus]
pir  A46199	nucleotide-exchange-factor homolog c...
pdb  1PTR	Molecule: Protein Kinase C Delta Ty...
gi  915330	(U24071) Munc13-2 [Rattus norvegicus]
gi  474982	(D21239) 'C3G protein' [Homo sapiens...
gi  1763306	(U75361) Munc13-3 [Rattus norvegicus]
gi  806957	guanine-nucleotide exchange factor C...
sp  Q03385 GNDS_MOUSE	GUANINE NUCLEOTIDE DISSOCIATION STIM...
pir  BVBYL1	LTE1 protein - yeast (Saccharomyces...
gi  452242	(D21354) a putative guanine nucleoti...
sp  P07866 LTE1_YEAST	LOW TEMPERATURE ESSENTIAL PROTEIN /P...
gi  509050	(Z22521) protein kinase C delta [Hom...
gi  520587	(D10495) protein kinase C delta-type...
sp  P05130 KPC1_DROME	PROTEIN KINASE C, BRAIN ISOZYME (PKC...
pir  S35704	protein kinase C (EC 2.7.1.-) delta...
sp  Q05655/KPCD_HUMAN	PROTEIN KINASE C, DELTA TYPE (NPKC-D...
pir  S40279	protein kinase C mu - human /pir A5...
sp  P09215 KPCD_RAT	PROTEIN KINASE C, DELTA TYPE (NPKC-D...
gi  520878	(Z34524) serine/threonine protein ki...
gi  1519719	(U68142) RalGDS-like [Homo sapiens]

FIG 12 (III)

32/85

FIG 12 IV

7.2e-13	1
3.4e-12	3
3.4e-12	3
5.5e-12	1
5.6e-12	1
1.5e-11	1
1.6e-11	2
3.3e-11	3
6.4e-11	2
7.8e-11	3
1.0e-10	2
1.9e-10	1
2.7e-10	1
2.7e-10	1
4.0e-10	1
4.6e-10	1
4.7e-10	1
4.7e-10	1
4.7e-10	1
4.9e-10	1
9.0e-10	1
1.8e-09	1
3.8e-09	3

157  
136  
136  
151  
149  
136  
150  
131  
153  
128  
133  
139  
139  
139  
137  
137  
137  
137  
137  
137  
135  
133  
115

33/85

FIG 13a (I)FIG 13a (II)FIG 13a (III)FIG 13a (IV)FIG 13a (V)FIG 13a (VI)FIG 13a (VII)FIG 13a (VIII)FIG 13a (IX)FIG 13a (X)FIG 13a

34/85

**FIGURE 13(a) (I)**

CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC CTC CCA AAA TAT	44
Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr	
1 5 10	
TCC CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG	89
Ser His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln	
15 20 25	
CTG TCC CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG	134
Leu Ser Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu	
30 35 40	
CCT GCA GGA AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA	179
Pro Ala Gly Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly	
45 50 55	
ACG CAG CGC CTG TGT GGC CGC CGG ACT CAA GGC TGG CCT GGC TCA	224
Thr Gln Arg Leu Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser	
60 65 70	
AGT GAA CAG CAC GTC CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT	269
Ser Glu Gln His Val Gln Glu Ala Thr Ser Ser Ala Gly Leu His	
75 80 85	

35/85

FIGURE 13(a) (II)															314
TCT	GGG	GTG	GAC	GAG	CTG	GGG	GTT	CGG	TCC	GAG	CCC	GGT	GGG	AGG	
Ser	Gly	Val	Asp	Glu	Leu	Gly	Val	Arg	Ser	Glu	Pro	Gly	Gly	Arg	
90					95					100					
CTC	CCG	GAG	CGC	AGC	CTG	GGC	CCA	GCC	CAC	CCC	GCG	CCG	GCG	GCC	359
Leu	Pro	Glu	Arg	Ser	Leu	Gly	Pro	Ala	His	Pro	Ala	Pro	Ala	Ala	
105					110					115					
ATG	GCA	GGC	ACC	CTG	GAC	CTG	GAC	AAG	GCG	TGC	ACG	GTG	GAG	GAG	404
Met	Ala	Gly	Thr	Leu	Asp	Leu	Asp	Lys	Gly	Cys	Thr	Val	Glu	Glu	
120					125					130					
CTG	CTC	CGC	GGG	TGC	ATC	GAA	GCC	TTC	GAT	GAC	TCC	GGG	AAG	GTG	449
Leu	Leu	Arg	Gly	Cys	Ile	Glu	Ala	Phe	Asp	Asp	Ser	Gly	Lys	Val	
135					140					145					
CGG	GAC	CCG	CAG	CTG	GTG	CGC	ATG	TTC	CTC	ATG	ATG	CAC	CCC	TGG	494
Arg	Asp	Pro	Gln	Leu	Val	Arg	Met	Phe	Leu	Met	Met	His	Pro	Trp	
150					155					160					
TAC	ATC	CCC	TCC	TCT	CAG	CTG	GCG	GCC	AAG	CTG	CTC	CAC	ATC	TAC	539
Tyr	Ile	Pro	Ser	Ser	Gln	Leu	Ala	Ala	Lys	Leu	Leu	His	Ile	Tyr	
165					170					175					



36/85

**FIGURE 13(a) (III)**

CAA	CAA	TCC	CGG	AAG	GAC	AAC	TCC	AAT	TCC	CTG	CAG	GTG	AAA	ACG	584
Gln	Gln	Ser	Arg	Lys	Asp	Asn	Ser	Asn	Ser	Leu	Gln	Val	Lys	Thr	
180					185					190					
TGC	CAC	CTG	GTC	AGG	TAC	TGG	ATC	TCC	GCC	TTC	CCA	GCG	GAG	TTT	629
Cys	His	Leu	Val	Arg	Tyr	Trp	Ile	Ser	Ala	Phe	Pro	Ala	Glu	Phe	
195					200					205					
GAC	TTG	AAC	CCG	GAG	TTG	GCT	GAG	CAG	ATC	AAG	GAG	CTG	AAG	GCT	674
Asp	Leu	Asn	Pro	Glu	Leu	Ala	Glu	Gln	Ile	Lys	Glu	Leu	Lys	Ala	
210					215					220					
CTG	CTA	GAC	CAA	GAA	GGG	AAC	CGA	CGG	CAC	AGC	AGC	CTA	ATC	GAC	719
Leu	Leu	Asp	Gln	Glu	Gly	Asn	Arg	Arg	His	Ser	Ser	Leu	Ile	Asp	
225					230					235					
ATA	GAC	AGC	GTC	CCT	ACC	TAC	AAG	TGG	AAG	CGG	CAG	GTG	ACT	CAG	764
Ile	Asp	Ser	Val	Pro	Thr	Tyr	Lys	Trp	Lys	Arg	Gln	Val	Thr	Gln	
240					245					250					
CGG	AAC	CCT	GTG	GGA	CAG	AAA	AAG	CGC	AAG	ATG	TCC	CTG	TTG	TTT	809
Arg	Asn	Pro	Val	Gly	Gln	Lys	Lys	Arg	Lys	Met	Ser	Leu	Leu	Phe	
255					260					265					

37/85

<b>FIGURE 13(a) (IV)</b>															854
GAC	CAC	CTG	GAG	CCC	ATG	GAG	CTG	GCG	GAG	CAT	CTC	ACC	TAC	TTG	
Asp	His	Leu	Glu	Pro	Met	Glu	Leu	Ala	Glu	His	Leu	Thr	Tyr	Leu	
270					275					280					
GAG	TAT	CGC	TCC	TTC	TGC	AAG	ATC	CTG	TTT	CAG	GAC	TAT	CAC	AGT	899
Glu	Tyr	Arg	Ser	Phe	Cys	Lys	Ile	Leu	Phe	Gln	Asp	Tyr	His	Ser	
285					290					295					
TTC	GTG	ACT	CAT	GGC	TGC	ACT	GTG	GAC	AAC	CCC	GTC	CTG	GAG	CGG	944
Phe	Val	Thr	His	Gly	Cys	Thr	Val	Asp	Asn	Pro	Val	Leu	Glu	Arg	
300					305					310					
TTC	ATC	TCC	CTC	TTC	AAC	AGC	GTC	TCA	CAG	TGG	GTG	CAG	CTC	ATG	989
Phe	Ile	Ser	Leu	Phe	Asn	Ser	Val	Ser	Gln	Trp	Val	Gln	Leu	Met	
315					320					325					
ATC	CTC	AGC	AAA	CCC	ACA	GCC	CCG	CAG	CGG	GCC	CTG	GTC	ATC	ACA	1034
Ile	Leu	Ser	Lys	Pro	Thr	Ala	Pro	Gln	Arg	Ala	Leu	Val	Ile	Thr	
330					335					340					

38/85

**FIGURE 13(a) (V)**

CAC TTT	GTC CAC	GTG GCG	GAG AAG	CTG CTA	CAG CTG	CAG AAC	TTC TTC	1079
His Phe	Val His	Val Ala	Glu Ala	Leu Leu	Gln Leu	Gln Asn	Phe	
345		350			355			
AAC ACG	CTG ATG	GCA GTG	GTC GGC	CTG GGC	CAC AGC	TCC ATC	ATC	1124
Asn Thr	Leu Met	Ala Val	Val Val	Gly Gly	His Ser	Ser Ser	Ile	
360		365			370			
TCC CGC	CTC AAG	GAG ACC	CAC AGC	CAC GTT	AGC CCT	GAG ACC	ATC	1169
Ser Arg	Leu Lys	Glu Thr	His Thr	Val Val	Ser Pro	Glu Thr	Ile	
375		380			385			
AAG CTC	TGG GAG	GGT CTC	ACG GAA	CTA GTG	ACG GCG	ACA GGC	AAC	1214
Lys Leu	Trp Glu	Gly Leu	Thr Thr	Val Leu	Thr Ala	Thr Gly	Asn	
390		395			400			
TAT GGC	AAC TAC	CGG CGT	CGG CTG	GCA GCC	TGT GTG	GGC TTC	CGC	1259
Tyr Gly	Asn Tyr	Arg Arg	Arg Leu	Ala Ala	Cys Val	Gly Phe	Arg	
405		410			415			
TTC CCG	ATC CTG	GGT GTG	CAC CTC	AAG GAC	CTG GTG	GCC CTG	CAG	1304
Phe Pro	Ile Leu	Gly Val	His His	Lys Asp	Leu Val	Ala Leu	Gln	
420		425			430			

39/85

**FIGURE 13(a) (VI)**

CTG GCA CTG CCT GAC	TGG CTG GAC CCA GCC CGG ACC CGG CTC AAC	1349
Leu Ala Leu Pro Asp	Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn	
435	440 445	
GGG GCC AAG ATG AAG CAG	CTC TTT AGC ATC CTG GAG GAG CTG GCC	1394
Gly Ala Lys Met Lys	Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala	
450	455 460	
ATG GTG ACC AGC CTG	CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG	1439
Met Val Thr Ser Leu	Arg Pro Pro Val Gln Ala Asn Pro Asp Leu	
465	470 475	
CTG AGC CTG CTC ACG	GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT	1484
Leu Ser Leu Leu Thr	Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp	
480	485 490	
GAG CTG TAC CAG CTG	TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC	1529
Glu Leu Tyr Gln Leu	Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser	
495	500 505	

40/85

**FIGURE 13(a) (VII)**

TCG CCA ACC AGC CCC	ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG	1574
Ser Pro Thr Ser Pro	Pro Pro Arg Pro Pro	
510	520	
GTA CTG GAG GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG	1619	
Val Leu Glu Glu Trp	Pro Lys Leu Asp Gln	
525	535	
GCC CTC GTG GTG GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC	1664	
Ala Leu Val Val Glu His	Val Glu Ser Val Phe	
540	550	
CGG AAC TTT GAC GTC GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA	1709	
Arg Asn Phe Asp Val	Ile Ser Gln Glu Glu	
555	565	
TTC CAG ATC ATC CGT GGG AAC TTC CCT TAC CTC AGC GCC TTT GGG	1754	
Phe Gln Ile Ile Arg	Leu Ser Ala Phe Gly	
570	580	
GAC CTC GAC CAG AAC CAG GAT GGC TGC ATC AGC AGG GAG GAG ATG	1799	
Asp Leu Asp Gln Asn	Arg Glu Glu Met	
585	595	

41/85

**FIGURE 13(a) (VIII)**

GTT TCC TAT TTC CTG	CGC	TCC	AGC	TCT	GTG	TTG	GGG	GGG	CGC	ATG	1844
Val Ser Tyr Phe Leu	Arg	Ser	Ser	Ser	Val	Leu	Gly	Gly	Arg	Met	
600	605					610					
GGC TTC GTA CAC AAC TTC	CAG	GAG	AGC	AAC	TCC	TTG	CGC	CCC	GTC	1889	
Gly Phe Val His Asn	Phe	Glu	Ser	Asn	Ser	Leu	Arg	Pro	Val		
615	620				625						
GCC TGC CGC CAC TGC AAA	GCC	CTG	ATC	CTG	GGC	ATC	TAC	AAG	CAG	1934	
Ala Cys Arg His Cys	Lys	Ala	Ile	Leu	Gly	Ile	Tyr	Lys	Gln		
630	635				640						
GGC CTC AAA TGC CGA GCC	TGT	GGA	GTG	AAC	TGC	CAC	AAG	CAG	TGC	1979	
Gly Leu Lys Cys Arg	Ala	Cys	Gly	Val	Cys	His	Lys	Gln	Cys		
645	650				655						
AAG GAT CGC CTG TCA GTT	GAG	TGT	CGG	CGC	AGG	GCC	CAG	AGT	GTG	2024	
Lys Asp Arg Leu Ser	Val	Glu	Cys	Arg	Arg	Ala	Gln	Ser	Val		
660	665				670						

42/85

**FIGURE 13(a) (IX)**

AGC CTG GAG GGG TCT GCA CCC TCA CCC ATG CAC AGC CAC	2069
Ser Leu Glu Gly Ser Ala Pro Ser Pro Met His Ser His	
675	685
CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG CGA	2114
His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg	
690	700
GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GTA CAG ACG GTG	2159
Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Val Gln Thr Val	
705	715
GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG	2198
Glu Asp Gly Val Phe Asp Ile His Leu *	
720	725
TGGTTGGATC AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA	2248
GCAGGGAGCC TGGGGGTGTC GGGGCAGGAG GCTGGGGATG GGGGTGGGAT	2293
ATGAGGGTGG CATGCAGCTG AGGCACGGC CAGGGCTGGT GTCCCTAAGG	2348

43/85

**FIGURE 13(a) (x)**

TTGTACAGAC	TCTTGTGAAT	ATTGTATTT	TCCAGATGGA	ATAAAAAGGC	2398
CCGTGTAATT	AACCTTC	(A)n			2416



44/85

**FIGURE 13(b)**

CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT 50  
 CTTGTCCCTAG CCATATCCCC AGACTATCTC AAGGACCAGC TGTCCTCCACG 100  
 CCCCCGACCT CCACTAGGCC TGTGCCACCC GCTGCCCTGCA GGAAGACGCC 150  
 CGGTCCCCGG CCGGGTTAG CCC CAT GGG AAC GGG GTT CGG TCC GAG 196

\* Pro His Gly Asn Gly Val Arg Ser Glu

1

5

CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC CCA GCC CAC 238

Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro Ala His

10

15

20

CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC CTG GAC AAG 280

Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys

25

30

35

GGC TGC ACG GTG GAG GAG CT 300

Gly Cys Thr Val Glu Glu Leu

40

45/85

FIG 14 (I)

FIG 14 (II)

FIG 14

46/85

FIGURE 14 (I)

1	MAGTLDLKGC...TVEELLRGCIEAF..DDSGKVDPQLVRMFLMMHPW	45
1	MSSKVEEDQHQLLTEDQLVARCVEC <u>FDVDEDEVEDIEFVDALFLSHQW</u>	50
46	YIPSSQLAAKLLHIYQQSRKDNSNSLQVKTC <u>HLVRYWISAFPAEFDLNPE</u>	95
51	LSDSLSLITHFVNFYQETRNVEQRE...AVCRAVSFWIEKFPMHFDAQPO	97
96	LAEQIKELKALLDQEGNRRHSSLIDIDSVPTYKKRQVTQORNVPVQKK..	143
98	VCAQVVRCLKTIAEDINENIRNGL.DVSALPFAWLRAVSVRNPLAKQTIV	146
144	.....RKMSLLFDHLEPMELAEHLTYLEYR	168
147	RVDFETLPTPGTPPPFPPIASKKFSLTAFSLSFVQASPSDISTSLSHIDYR	196
169	SFCKILFQDYHSFVTHGCTVDNPVLERFISLFNSVSQWVQMLSKPTAP	218
197	VLSTISITELKQYVKDGHRLRSCPMLEERSISVFNNLSNWVQCLILNKTTPK	246
219	QRALVITHEVHVAAEKLLOLONFNTLMVAVGGLSHSSISRLKETHSHVSPE	268
247	ERAEILVKFVHVAAKHLRKINNNTLMVAVGGLSHSSISRLKETHSHVSPE	296
269	TIKLWEGLTELVTATGNYGNRYRRRLAAC.VGFRFPILGVHLKDLVALQLA	317
297	IKKELTQLTNLLSAQHNFCYRKALGACNKKFRIPPIIGVHLKDLVAINCS	346

47/85

## FIGURE 14 (II)

318	LPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPV.QANPDLLSLLTV	366
347	GANFEKT..KCISSDKLVKLSKLLSNFLVFNQKGHNLPENMDLINTLKV	394
367	SLDQYQTEDELYQLSLOREPRSKSSPTSCTPPPRPPVLEEWTSAAKP	416
395	SLDIRYNDDDIYELSLRREP KTFMN.....FEPSRGLVFAEWASGVTV	437
417	KLDQALVVEHIEKMVESVFRNFDVDGDGHI <del>SOEEFQ</del> IIRGNFPYLSAFGD	466
438	APDNA TVSKHISAMVDAVFKHYDHD <del>RDGFISOE</del> EFQLIAGNFPFIDAFVN	487
467	L <del>DONODGCISREEMVSYFLRSS.SVLGGRMGFVHNFOESNSLRPVACRHC</del>	515
488	<del>IDVDMDGOISKDELKTYFMAANKNTKDLRRGFKHNFHETTELPTTCNHC</del>	537
516	KALILGIYKOGKCRACGVNCHKOCKDRLSVE <del>CRRAQ</del> SVSLEGSAPSPS	565
538	NKLLWGILROGFKCKDCGLAVHSCCKSNAVAECRRKSSSNLTRA <del>AEWFAS</del>	587
566	PMHSHHHRAFSFSLPRPGRGRSRPPPEIREEEVQTVEDGVFDIHL	609
588	PRGSMRSRIINTC...NNSGSTPDEEIGLVSLACEEVFEDDDL	627

48/85

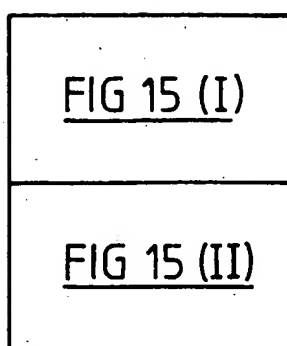


FIG 15

**FIGURE 15 (I)**

human	CGATTTCATT	CCTCGCTCCC	CACAGGTCCC	TCTCCCCAAA	ATATTCCCAT	CTTGTCCTAG	60
human	CCCATCCCCC	AGACTATCTC	AAGGACCAGC	TGTCCCCACG	CCCCCGACCT	CCACTAGGCC	120
human	TGTGCCACCC	GCTGCCCTGCA	GGAAGACGCC	CGTCCCCGGG	CCGGGTAGC	CCCATGGGAA	180
human	CGCAGCGCCT	GTGTGGCCCG	GGGACTCAAG	GCTGGCCTGG	CTCAAGTGAA	CAGCACGTCC	240
mouse			***tcag**	***ag***	t*****	***a*g***t>	
human	AGGAGGCGAC	CTCGTCCGCG	GGTTTGCAAT	CTGGGGTGGA	CGAGCTGGGG	GTTCGGTCCG	300

acagg

mouse	g*****t**a	**-*catt**	*****	***aa**aa*	g*ct*****	**a*aat**>	
human	AGCCCGGTGG	GAGGCTCCCG	GAGCGCAGCC	TGGGCCCCAGC	CCACCCCGCG	CCGGCGGCCA	360
mouse	***a*t****	*****tga	***t*t*a*t	***t*t***	***-tg**a	*****a****>	
human	TGGCAGGCAC	CCTGGACCTG	GACAAGGGCT	GCACGGTGGA	GGAGCTGCTC	CGCGGGTGCA	420
mouse	***ga****	t*****	*****t*	***c*****	*****	**t**c***t**>	
human	TCGAAGCCTT	CGATGACTCC	GGGAAGGTGC	GGGACCCGCA	GCTGGTGCGC	ATGTTCTCTCA	480

49/85

**FIGURE 15 (II)**

```

mouse *****t*****a*****t*****>
human TGATGCACCC CTGGTACATC CCTCCTCTC AGCTGGCGGC CAAGCTGCTC CACATCTACC 540
mouse *****a*****t*****g*****t*****>
human AACAAATCCCG GAAGGACAAC TCCAATTCCC TGCAGGTGAA AACGTGCCAC CTGGTCAGGT 600
mouse *g*****a*****t*****t*****>
human ACTGGATCTC CGCCTTCCCA GCGGAGTTG ACTTGAACCC GGAGTTGGCT GAGCAGATCA 660
mouse *****a*****c*****a*****>
human AGGAGCTGAA GGCTCTGCTA GACCAAGAAG GGAACCGACG GCACAGCAGC CTAATCGACA 720
mouse *****t*****c*****>
human TAGACAGCGT
mouse *c**g**t**

```

51/85

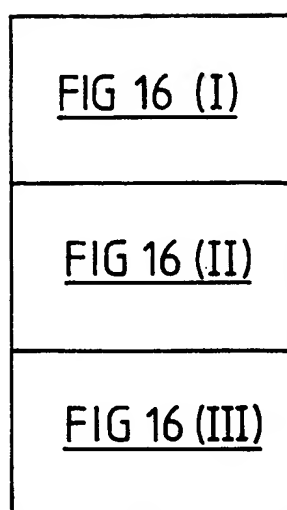


FIG 16



52/85

**FIGURE 16 (I)**

CACGCCCTCGG	AAGGAGGTT	TGGGTCGGT	GGTTACACAG	TGAGTGTC	50
TGAAGCCAAA	TGGTCGGAAA	CCGTTACCCG	CTCTCCTAG	GCC CGG CTA	98
				* Ala Arg Leu	
				1	
GTG GGG ACC CCA ACC GCC TGC GGC TGC CCC TCC CAA GTT CCT					140
Val Gly Thr Pro Thr Ala Cys Gly Cys Pro Ser Gln Val Pro					
				15	
				10	
				5	
CCC TGT TGG CCA CCG CCA CAT GCG GCT GCC GCT GCG					182
Pro Cys Trp Pro Gly Ile Gln Val Ser Ser Leu Arg Ala Ala					
				25	
				20	
GAG AAC CCA CCG CCA CAT GCG GCT GCC CCT TTC CAT TCG ACC					224
Glu Asn Pro Pro Pro His Ala Ala Ala Pro Phe His Ser Thr					
				35	
				40	
				45	
CTG TGG GGA GCC AGG CTT CCG GGG CCC CGT TCC TCC TGT GTG					266
Leu Trp Gly Ala Arg Leu Pro Gly Pro Arg Ser Ser Cys Val					
				50	
				55	
AAC TGG GCC CCC CGC CCC CAT TCC CAG ACA TCA AGG CCG CGT					308
Asn Trp Ala Pro Arg Pro His Ser Gln Thr Ser Arg Pro Arg					
				60	
				65	
				70	

53/85

**FIGURE 16 (II)**

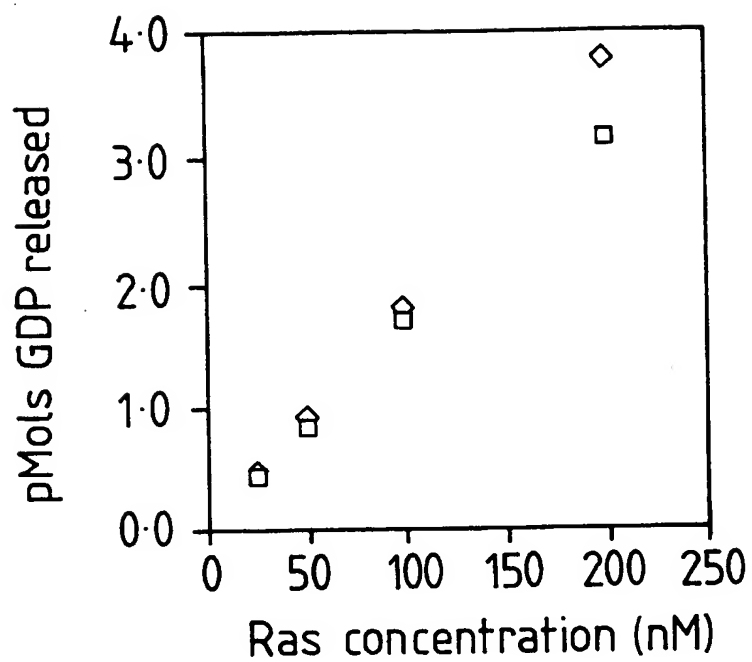
CTC	CAG	ATA	GCC	ACG	ATT	TCA	TTC	CTC	GCT	CCC	CAC	AGG	TCC	350
Leu	Gln	Ile	Ala	Thr	Ile	Ser	Phe	Leu	Ala	Pro	His	Arg	Ser	
75						80					85			
CTC	TCC	CCA	AAA	TAT	TCC	CAT	CTT	GTC	CTA	GCC	CAT	CCC	CCA	392
Leu	Ser	Pro	Lys	Tyr	Ser	His	Leu	Val	Leu	Ala	His	Pro	Pro	
90							95					100		
GAC	TAT	CTC	AAG	GAC	CAG	CTG	TCC	CCA	CGC	CCC	CGA	CCT	CCA	434
Asp	Tyr	Leu	Lys	Asp	Gln	Leu	Ser	Pro	Arg	Pro	Arg	Pro	Pro	
105							110						115	
CTA	GGC	CTG	TGC	CAC	CCG	CTG	CCT	GCA	GGA	AGA	CGC	CCG	GTC	476
Leu	Gly	Leu	Cys	His	Pro	Leu	Pro	Ala	Gly	Arg	Arg	Pro	Val	
120									125					
CCG	GGC	CGG	GTT	AGC	CCC	ATG	GGA	ACG	CAG	CGC	CTG	TGT	GGC	518
				*	Pro	His	Gly	Asn						
Pro	Gly	Arg	Val	Ser	Pro	Met	Gly	Thr	Gln	Arg	Leu	Cys	Gly	
130					135					140				

54/85

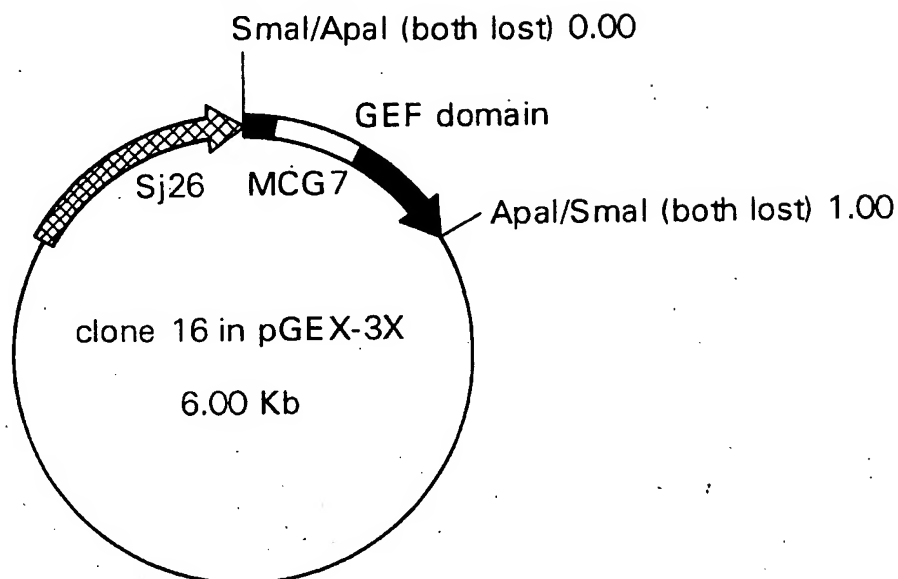
**FIGURE 16 (III)**

CGC GGG ACT CAA	GGC TGG CCT GGC TCA AGT GAA CAG CAC GTC	560
Arg Gly Thr Gln Gly Trp	Pro Gly Ser Ser Glu Gln His Val	
145	150	155
CAG GAG GCG ACC TCG TCC TCG GGT TTG CAT TCT GGG GTG GAC		602
Gln Glu Ala Thr 155	Ala Gly Leu His Ser Gly Val Asp	
160	165	
GAG CTG GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG		644
Glu Leu Gly Val 170	Glu Pro Gly Arg Gly Pro Glu	
175	180	
CGC AGC CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA		686
Arg Ser Leu Gly 185	Pro Ala Pro Ala Met Ala	
190		
GGC ACC CTG GAC CTG GAC AAG GGC TGC ACG GTG G		720
Gly Thr Leu Asp 195	Lys Gly Cys Thr Val	
200	205	

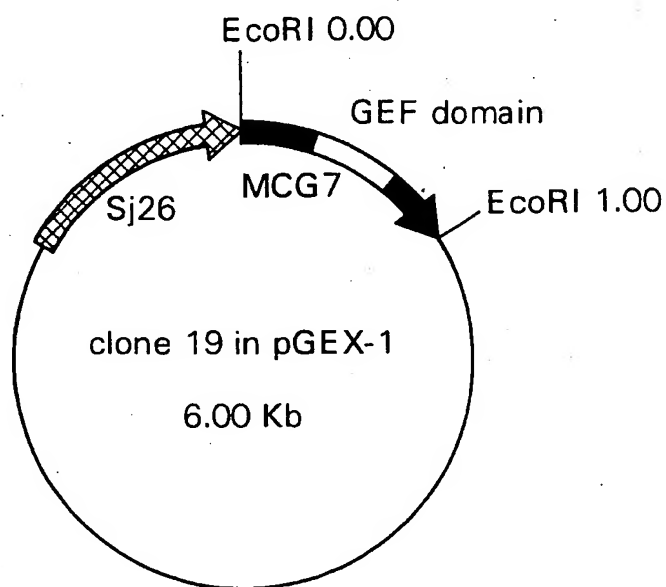
55/85

FIGURE 17

56/85

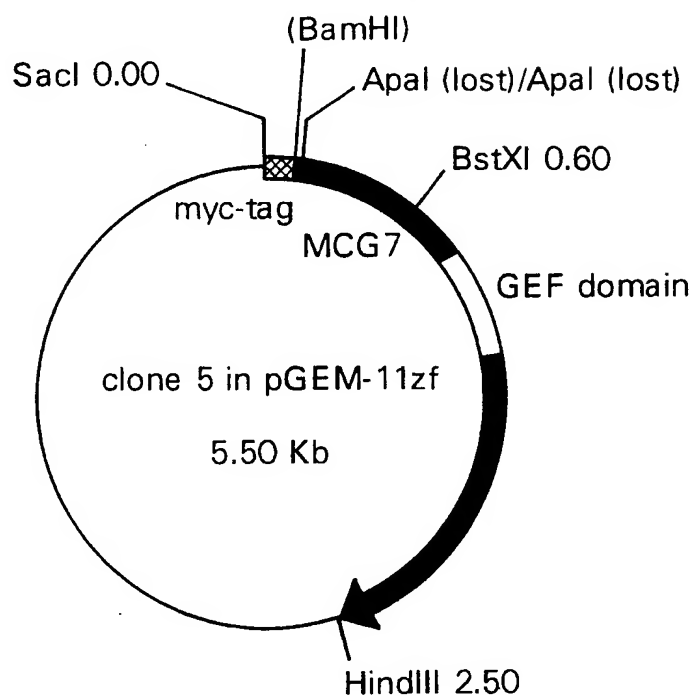
FIGURE 18 (Cont. I)

Plasmid name: clone 16 in pGEX-3X  
Plasmid size: 6.00 kb

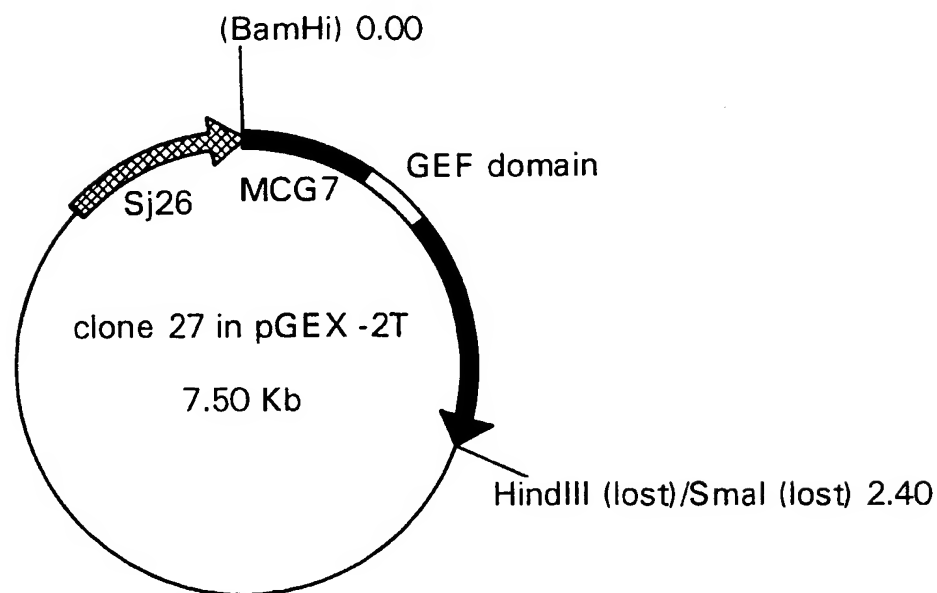
FIGURE 18 (Cont. II)

Plasmid name: clone 19 in pGEX-1  
Plasmid size: 6.00 Kb

57/85

FIGURE 18 (Cont. III)

Plasmid name: clone 5 in pGEM-11zf  
Plasmid size: 5.50 kb



Plasmid name: clone 27 in pGEX-2T  
Plasmid size: 7.50 kb

FIGURE 18 (Cont. IV)

SUBSTITUTE SHEET (RULE 26)

58/85

<u>FIG 19 (I)</u>
<u>FIG 19 (II)</u>
<u>FIG 19 (III)</u>
<u>FIG 19 (IV)</u>

FIG 19

59/85

**FIGURE 19 (I)**

43	GGCCGGCCGCC	ATG	CCG	CCC	TTA	CTG	CCC	CTG	CGC	CTG	TGC	CGG
	Met	Pro	Pro	Leu	Leu	Pro	Pro	Leu	Arg	Leu	Cys	Arg
		1			5						10	
85	CTG	TGG	CCC	AAC	CCT	CCC	TCC	CGG	CTC	CTC	GGA	GCG
	Leu	Trp	Pro	Arg	Asn	Pro	Pro	Ser	Arg	Leu	Gly	Ala
			15					20				25
127	GCC	GGG	CAG	CGG	TCC	AGA	CCC	AGT	ACT	TAT	GAA	CTG
	Ala	Gly	Gln	Arg	Ser	Arg	Pro	Ser	Thr	Tyr	Glu	Leu
				30						35		
169	GGG	GTG	CAT	CCT	GGT	GCC	AGC	ACT	GAG	GAA	GTT	AAA
	Gly	Val	His	Pro	Gly	Ala	Ser	Thr	Glu	Val	Lys	Arg
					40	45				50		Ala
211	TTC	TTC	TCC	AAG	TCC	AAA	GAG	CTG	CAC	CCA	GAC	GAC
	Phe	Phe	Ser	Lys	Ser	Lys	Glu	Leu	His	Pro	Asp	Arg
				55		60					65	Pro
253	GGG	AAC	CCA	AGC	CTG	CAC	AGC	CGC	TTT	GTG	GAG	CTG
	Gly	Asn	Pro	Ser	Leu	His	Ser	Arg	Phe	Val	Glu	Leu
				70				75				80



60/85

**FIGURE 19 (II)**

GCA TAC CGT GTG	CTC AGC CGT GAG CAG AGC CGC CGC AGC TAT	295
Ala Tyr Arg Val	Leu Ser Arg Glu Gln Ser Arg Arg Ser Tyr	85 90 95
GAT GAC CAG CTC	CGC TCA GGT AGT CCC CCA AAG TCT CCA CGA	337
Asp Asp Gln Leu	Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg	100 105
ACC ACA GTC CAT	GAC AAG TCT GCC CAC CAA ACA CAC AGC TCC	379
Thr Thr Val His	Asp Lys Ser Ala His Gln Thr His Ser Ser	110 115 120
TGG ACA CCC CCC	AAC GCA CAG TAC TGG TCC CAG TTT CAC AGC	421
Trp Thr Pro Pro	Pro Asn Ala Gln Tyr Trp Ser Gln Phe His Ser	125 130 135
GTG AGG CCA CAG	GGG CCC CAG TTG AGG CAG CAA CAC AAA	463
Val Arg Pro Gln	Pro Gly Pro Gln Leu Arg Gln Gln His Lys	140 145 150
CAA AAC AAA CAA	GTG CTG GGG TAC TGC CTC CTC ATG CTG	505
Gln Asn Lys Gln	Val Leu Gly Tyr Cys Leu Leu Met Leu	155 160 165



62/85

**FIGURE 19 (IV)**

813	CGGTCCCCG	TTTGCTTCCT	TCCCTGGACG	GCCCGCTCCC	CGAAACGCGC
832	GCAATAAAGT	GATTCGCAG			

**FIGURE 20**

>sp|P08622|DNAJ\_ECOLI DNAJ PROTEIN >pir ||HHECDJ heat shock protein

dnaJ -

Escherichia coli >gi |145769 (M12565)-heat shock protein dnaJ

[Escherichia coli] >gi |216441 (D10483) dnaJ protein

[Escherichia coli]

Length = 376

Score = 138 (63.7 bits), Expect = 1.2e-10, P = 1.2e-10

Identities = 25/62 (40%), Positives = 39/62 (62%)

63/85

Query: 35 YYELLGVHPGASTEVEVKRAFFSKSKELHPDRDPGNPSLHSRFFVELSEAYRVLSREQSRRS94

YYE+LGV A E+++A+ + + HPDR+ G+ ++F E+ EAY VL+ Q R +

Sbjct: 6 YYEILGVSKTAEEREIRKAYKRLAMKYHPDRNQGDKEAEAKFKEIKEAYEVLTDTSQKRAA65

Query: 95 YD 96

YD

Sbjct: 66 YD 67

64/85

<u>FIG 21 (I)</u>
<u>FIG 21 (II)</u>
<u>FIG 21 (III)</u>
<u>FIG 21 (IV)</u>

FIG 21

65/85

**FIGURE 21 (I)**

>gi|1703590 (U80439) contains similarity to a DNAJ-like domain  
(Caenorhabditis elegans]

Length = 345

Score = 98 (45.2 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12  
Identities = 17/37 (45%), Positives = 28/37 (75%)

Query: 28 QRSRPSTYYELLGVHPGASTE EVKRAFFSKSKELHPD 64

++ R T+YE+LGV A+ E+K AF+++SK++HPD

Sbjct: 22 KKIRQRTHEYEV LGVESTATLSEIKSAFYAQSKKVHPD 58

Score = 74 (34.1 bits), Expect: = 5.2e-12, Sum P(3) = 5.2e-12  
Identities = 17/32 (53%), Positives = 19/32 (59%)

66/85

**FIGURE 21 (II)**

Query: 71 SLHSRFVELSEAYRVLSREQSRRSYDDQLRSG 102

S + F+EL AY VL R RR YD QLR G

Sbjct: 64 SATASFLELKNAYDVLRRPADRRRLYDYQLRGG 95

Score = 39 (18.0 bits), Expect =  $5.2e-12$ , Sum P(3) =  $5.2e-12$   
Identities = 10/42 (23%), Positives = 19/42 (45%)

Query: 162 LLMLAGMGLHYIAFRKVKQMHLNFMDEKDRITAFYNEARAR 203

L+++AG Y+ Q L+ + ++D I F + R

Sbjct: 158 LVLVAGYNGGYLYLLAYNQQLDKLIDEDEDIAKCFRLRQKEFR 199

67/85

**FIGURE 21 (III)**

>gnl |PID |e281266 (Z81030) COLG10.12 [Caenorhabditis elegans]

Length = 191

Score = 96 (44.3 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09

Identities = 17/41 (41%), Positives = 27/41 (65%)

Query: 35 YYELLGVHPGASTE EVKRAFFSKSKELHPDRDPGNPSLHSR 75

YYE++GV A+ +E++ AF K+K+LHPD+ + SR

Sbjct: 19 YYEIIGVSASATRQEIRDAFLKKTKQLHPDQSRKSSKSDSR 59

Score = 54 (24.9 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09

Identities = 10/22 (45%), Positives = 15/22 (68%)

Query: 75 RFVELSEAYRVLSREQSRRSYD 96

+F+ + EAY VL E+ R+ YD



68/85

**FIGURE 21 (IV)**

Sbjct: 71 QFMLVKEAYDVLNRNEEKREYD 92

Score = 35 (16.1 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09  
Identities = 9/44 (20%), Positives = 22/44 (50%)

Query: 141 QGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFRKVKQMHLN 184

+ P+ + KQ ++L ++A +G + + RK++ L+

Sbjct: 145 RNPDEYLRKQKNRMLVVLAAATVMALIGANIVYIRKLQADRLS 188

69/85

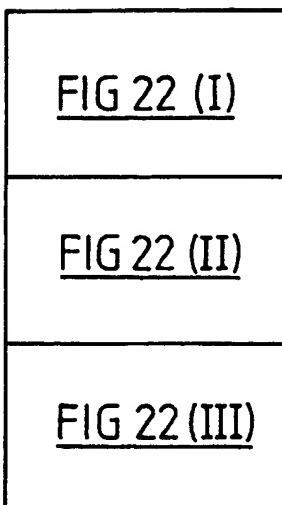


FIG 22

70/85

**FIGURE 22 (I)**

>sp|Q10209|YAY1\_SCHPO HYPOTHETICAL 44.8 KD PROTEIN C4H3.01 IN  
 CHROMOSOME I  
 >gi|1184014 (Z69380) unknown [Schizosaccharomyces pombe]

Length = 392

Score = 84 (38.8 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08  
 Identities = 13/36 (36%), Positives = 25/36 (69%)

Query: 35 YEELLGVHPGASTE EVKRAFFSKSKELHPDRDPGNP 70

YY+LLG+ A+ ++K+A+ + + HPD++P +P

Sbjct: 9 YYDLLGISTDATAVDIKKAYRKLAVKYHPDKNPDDP 44

Score = 64 (29.5 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08  
 Identities = 14/40 (35%), Positives = 23/40 (57%)

71/85

**FIGURE 22 (II)**

Query: 75 RFVELSEAYRVLSREQSRRSYDDQLRSGPPKSPRTTVHD 114

+F ++SEAY+VL E+ R YD + + P+ T +D

Sbjct: 50 KFKISEAYQVLGDEKLRSQYDQFGKEKAVPEQGFDDAYD 89

Score = 37 (17.1 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08  
Identities = 9/29 (31%), Positives = 15/29 (51%)

Query: 190 DRIITAFYNEARARARANRGILQQERQRL 218

DR A E A A+ + +++ RQR+

Sbjct: 149 DRKKNAQIREREALAKREQEMIEDRRQRI 177

Score = 33 (15.2 bits), Expect = 0.00081, Sum P(3) = 0.00081

Identities = 8/19 (42%), Positives = 11/19 (57%)

72/85

**FIGURE 22 (III)**

Query: 140 PQGPQLRQQQHKQNKQVLG 158

PQG + Q+ + QVLG

Sbjct: 44 PQGASEKFQKISEAYQVLG 62

**FIGURE 23**

>gnl|PID|e253406 (X77635) tumorous imaginal discs [Drosophila virilis]  
 >gnl|PID|e263866 (Y07700) Tid58 protein [Drosophila virilis]  
 Length = 529

Score = 153 (70.6 bits), Expect = 9.7e-13, P = 9.7e-13  
 Identities = 27/71 (38%), Positives = 44/71 (61%)

73/85

Query: 26

AGQSRPSTYYELLGVHPGASTE EVKRAFFSKSKELHPDRDPGNPSLHSRFFVELSEAYRV 85  
 + R + YY LGV A+ +++K+A++ +K+ HPD + +P +F ++SEAY V

Sbjct: 72

SSSRMQAKDYATLGVAKNANAKDIKKAYYELAKKYHPDTNKDDPDASKKFQDVSEAYEV 131

Query: 86 LSREQSRRSYD 96

LS +Q RR YD

Sbjct:132 LSDDQKRREYD 142

74/85

FIG 24 (I)

FIG 24 (II)

FIG 24 (III)

FIG 24

75/85

FIGURE 24 (I)

MCG18	-----MPPLLPLRLCRLWP-RN--PP-----SRLLGAA
HDJ-2	MVKETTYDVLGVKPNATQEELKKAYRKLALKYHPDKN--PN-----EGEKFKQISQAYEV
HDJ-1	MGKD--YYQTLGLARGASDEEIKRAYRRQALRYHPDKNKEPG---AEEKFKEIAEAYDV
HSJ1	M-AS--YYEILDVPRASADDIKKAYRRKALQWHPDKN--PDNKEFAEKKFKEVAEAYEV
	. . . * * *
MCG18	AGQSRPSTY--YELLGVH-----PGA-----ST-EEVKRAFFS-
HDJ-2	LSDAKKRELYDKGGEQAIAK-----EGGAGG-----FGSPMDIFDMFFGGG
HDJ-1	LSDPRKREIFDRYGEEGLKGSGP-----SGSGGGANGTSFSYTFHGDPHAMFAEFFG--
HSJ1	LSDKHKREIYDRYGREGLTGTGTPSRAEAGSGGP--G--FTFT-FRSPEEVFREFFG--
	. . . * . **
MCG18	KSKELHPDRDPGNP-----SLHSRFVELSEAYRVLRSREQSRRS--YDDQLRSGSPPKSPRT
HDJ-2	GRMQRERRGKNVVHQLSVTLEDLYNGATRKLALQKNVICDKCEGRGGKKGAVECCPNCRG
HDJ-1	GRNPFDTFFGQRNGEEGMDIDDPFSGFPMGMGGFTNVNFGRS--RSAQEPARKKQDPPVT
HSJ1	SGDPFAELFDDDLGP--FSELQNRGSRHSGPFFTFSSSPGHSDFSSSSFSFSPGAGAFRS
	. . . .



76/85

## FIGURE 24 (II)

MCG18	TVHDKSAHQTHSSWTPPNAQY---WSQFHSVRPQ-----GP-----QLRQQQHKQN
HDJ-2	TGMQIRIHQIGPGMVQIQSVCMECQGHGERISPK-DRCKSCNGRKIVREKKILEVHIDK
HDJ-1	HDLRVSLEEIYSGCTKKMK-----ISH-KRLNP---D-----GKSIRNEDKILTIEVKK
HSJ-1	VSTSTTFVQGRRIITRRIME-----NGQ-ERVEVEED-----GQ-----LKSVTINGVPD * . . . *
MCG18	KQVLGYCLLL-----MLAGMGLHYIAFRKVKQMHLNFMDE-KDRIITAFYNEARARAN
HDJ-2	GMKDGQKITFHGEGDQEPGLEPGDIIIVLDQKDHAVFTRRGEDLFMCMDIQLVEALCGFQ
HDJ-1	GWKEGTKITFPKEGDQTSNNIPADIVFVLKDKPHNIFKRDGSDVIYPARISLREALCGCT
HSJ1	DLARGLELSR-RE--QQP-SVTSRSGGTQVQQTASCPLD-SDLSEDEDLQLAMAYSLSE * . . *
MCG18	RGILQQERQRLGQRQPP-PSEPTQGPEIVPRGAGP-----
HDJ-2	KPISTLDNRTIVITSHPGQIVKHGDIKCVLNEGMPYRRPYEKGRLLIEFKVNFENGFL
HDJ-1	VNVPTLDGRTIPVFK--DVIRPGMRKVPGEGLPLPKTPEKRGDLIIIEFEVIFPER--I
HSJ1	MEAAGKKPAGGREAQHR-RQGRPRPSTKIQAWGGP--RR--VRG--VKQPNVHPQR-RR * . *

77/85

FIGURE 24 (III)

MCG18	-----
HDJ-2	SPDKLSLLEKLLPERKEVEETDEMDDQVELVDFDPNQERRRRHYNGEAYEDDDEHHPRGGVQC
HDJ-1	PQTSRTVLEQVLPI
HSJ1	PLAASSEHRAQPD-----LIQILTGGSDSLWEEKRGVS-----

MCG18	---
HDJ-2	QTS
HDJ-1	---
HSJ1	---

\* = amino acid identity in all 4 proteins

- = conservative substitution

78/85

<u>FIG 25 (I)</u>
<u>FIG 25 (II)</u>
<u>FIG 25(III)</u>
<u>FIG 25 (IV)</u>

FIG 25

79/85

**FIGURE 25 (I)**

CAAGGAGCCT	CTGCCTGCCC	GTCGTCGTC	ATG	CCG	TCC	CTG	TTG	CTC	47
			Met	Pro	Ser	Leu	Leu	Leu	
			1				5		
CAG CTG CCC	CTG CGC	CTA TGC	CGG CTG	TGG CCG	CAT AGC	CTT			89
Gln Leu Pro	Leu Arg	Leu Cys	Arg Leu	Trp Pro	His Ser	Leu			
	10		15			20			
TCC ATC CGA	CTT CTC	ACA GCC	ACC ACA	GGG CAG	CGG TCT	GTC			131
Ser Ile Arg	Leu Leu	Thr Ala	Thr Thr	Gly Gln	Arg Ser	Val			
		25	30						
CCT ACT AAT	TAC TAT	GAA TTG	GGC GTG	CAT CCG	GGT GCC				173
Pro Thr Asn	Tyr Tyr	Glu Leu	Gly Val	His Pro	Gly Ala				
		40	45						
AGC GCT GAA	GAG ATT	AAA CGT	TTT TTC	ACC AAG	TCA AAA				215
Ser Ala Glu	Glu Ile	Lys Arg	Phe Phe	Thr Thr	Lys Ser	Lys			
		55			60				
GAG CTA CAC	CCT GAT	CGA GAC	GGG AAC	CCA GCC	CTG CAT				257
Glu Leu His	Pro Asp	Arg Asp	Gly Asn	Pro Ala	Leu His				
		65	70		75				

80/85

**FIGURE 25 (II)**

AGC CGC TTT GTG	GAG CTG AAT GAG GCA TAT CGA GTG CTC AGT	299
Ser Arg Phe Val	Glu Leu Asn Glu Ala Tyr Arg Val Leu Ser	80 85 90
CGT GAG GAA AGT	CGT CGT AAC TAT GAC CAC CAG CTG CAT TCA	341
Arg Glu Glu Ser	Arg Arg Asn Tyr Asp His Gln Leu His Ser	95 100
GCC AGT CCT CCA AAG	TCT TCA GGG AGC ACA GCC GAG CCT AAG	383
Ala Ser Pro Pro	Lys Ser Ser Thr Ala Glu Pro Lys	105 110 115
TAT ACG CAA CAG	ACA CAC AGC AGC TCC TGG GAA CCC CCC AAC	425
Tyr Thr Gln Gln	Thr Thr His Ser Ser Ser Trp Glu Pro Pro Asn	120 125 130
GCT CAA TAC TGG	GCC GCG TTT CAG TTC CAC AGT GTG AGG CCG CAG GGG	467
Ala Gln Tyr Trp	Ala Ala Trp Phe His Ser Val Arg Pro Gln Gly	135 140 145
CCG GAG TCA AGG	AAG AAG CAG CAG AAA CAC AAC CAG CGG GTC	509
Pro Glu Ser Arg	Lys Lys Gln Gln Arg Lys His Asn Gln Arg Val	150 155 160

81/85

FIGURE 25 (III)											
CTG GGG TAC TGC	CTC	CTG	CTC	ATG	GTG	GCA	GGC	ATG	GGC	CTG	551
Leu Gly Tyr Cys	Leu	Leu	Leu	Met	Val	Ala	Gly	Met	Gly	Leu	
	165					170					
CAC TAT GTT GCC	TTC	AGG	AAG	CTG	GAG	CAG	GTG	CAT	CGC	AGC	593
His Tyr Val Ala Phe	Arg	Arg	Lys	Leu	Glu	Gln	Val	His	Arg	Ser	
175	180					185					
TTC ATG GAT GAA AAG	GAC	GAC	CGG	ATC	ATT	ACA	GCC	ATC	TAC	AAT	635
Phe Met Asp Glu Lys	Asp	Arg	Arg	Ile	Ile	Thr	Ala	Ile	Tyr	Asn	
190	195					200					
GAC ACT CGG GCC	AGG	GCC	AGG	GCC	AAC	AGA	GCC	AGG	ATT	CAG	677
Asp Thr Arg Ala Arg	Ala	Ala	Arg	Ala	Asn	Arg	Ala	Arg	Ile	Gln	
205	210					215					
CAG GAG CGC CAC	GAG	AGG	CAG	CAG	CCT	CGG	GCA	GAA	CCC	TCC	719
Gln Glu Arg His	Glu	Arg	Gln	Gln	Pro	Arg	Ala	Glu	Pro	Ser	
220	225									230	

82/85

**FIGURE 25 (IV)**

CTG CCT CCA GAA AGC TCC AGG ATC ATG CCC CAG GAC ACA AGC 761  
 Leu Pro Pro Glu Ser Ser Arg Ile Met Pro Gln Asp Thr Ser  
 235 240

CCC TGAGAGGCTT AACTAAATGG GACCTTCATT GGTCCTCTCC CTGCTGCCCTG 814  
 Pro \*  
 245

TCCAGAACTA CACGTGCAAT AAATCATTT TCAG (A)n 849

83/85

FIGURE 26

```

human MCG18 MPPLL---PLRLCRLWPRNPSPRLGAAAGQSRSPSTYYELLGVHPGASTE EVKRAFFSK
mouse MCG18 MPSLLLQLPLRLCRLWPHSLSIRLLTAATGQRSVPTNYVELLGVHPGASAE EIKRAFFTK
**          *****          *** **          *...*****          *****.*
**          *****          *** **          *...*****          *****.*

human MCG18 SKELHPDRDPGNPSLHSRFVELSEAYRVL SREQSRRSYDDQLRSGSPPKSPRTTVHDKSA
mouse MCG18 SKELHPDRDPGNPALHSRFVELNEAYRVL SREESRRNYDQLHSASPPKSSGSTAEPKYT
*****          *****          *****          *****          *****.*
*****          *****          *****          *****          *****.*

human MCG18 HQTHSS-WTPPNAQYWSQFHSVRPQGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFR
mouse MCG18 QQTHSSWEPPNAQYWAQFHSVRPQGPESRKQQRKHNQRVLG YCLLLMVAGMGLHYVAFR
*****          *****          *****          *****          *****.*
*****          *****          *****          *****          *****.*

human MCG18 KVKQMHLNFMDEKDRIITAFYNEARARANRGILQQERQRLGQRQPPPEPTQGPE---
mouse MCG18 KLEQVHRFSFMDEKDRIITAIYNDTRARANRARIQQR---HERQQPRAEPSLPPESSR
*.*.*          *****          *****          *****          *****.*
*.*.*          *****          *****          *****          *****.*

human MCG18 IVPRGAGP
mouse MCG18 IMPQDTSP
*.*.*          *

```



84/85

## FIGURE 27

TTGA AGT CTA GCC CCA TCC TGG TCC AAT GCG CTC TTG GTA	40
* Ser Leu Ala Pro Ser Trp Ser Asn Ala Leu Leu Val	
1 5 10	
GCC TCC TTT CCC AGC TGC CCG CCC GCC GCG ATG CCG CCC TTA	82
Ala Ser Phe Pro Ser Cys Pro Pro Ala Ala Met Pro Pro Leu	
15 20 25	
CTG CCC CTG CGC CTG TGC CCG CTG TGG CCC CGC AAC CC	120
Leu Pro Leu Arg Leu Cys Arg Leu Trp Pro Arg Asn Pro	
30 35	

85/85

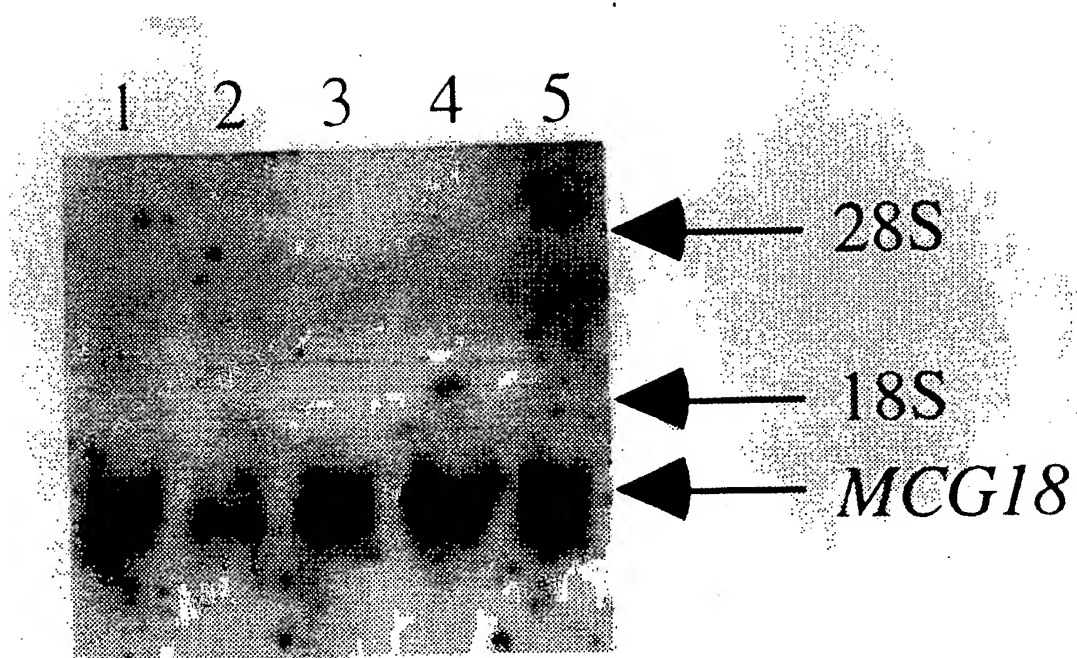
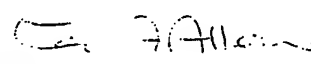


FIG 28

# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/AU 98/00380

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
Int Cl <sup>6</sup> : C12N 15/12; C07K 14/47; C07K 16/18; G01N 33/53		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) I/C: WPAT (D gene) Sequences provided by Applicant		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EMBL, Genbank, Swiss Prot and PIR: Sequences provided by applicant		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	Kedra D, Seroussi E, Fransson I, Trifunovic J, Clark M, Lagercranz J, Blennow E, Mehlin H, Dumanski J, Human Genetics, October 1997 100(5-6) 611-619 The germinal centre kinase gene and a novel CDC25-like gene are located in the vicinity of the PYGM gene on 11q13 EMBL AC Y12339	1-3,8-10,15-18
P,X	Guru S C, Agarwal S K, Manickain P, Olufemi S E, et al Genome Research, July 1997 7(7) 725-735. A transcript map for the 2.8-Mb region containing the multiple endocrine neoplasia type I locus TREMBL AC 014616	1, 4-5, 8, 11-12, 15, 19-21
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C <input type="checkbox"/> See patent family annex		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "I" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 16 July 1998		Date of mailing of the international search report 20 JUL 1998
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929		Authorized officer GILLIAN ALLEN  Telephone No.: (02) 6283 2266

# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/AU 98/00380

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 1, 2, 4, 6  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
They are to known groups of proteins and lack distinguishing features which would enable a meaningful search.
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Invention 1, defined by claims 2, 3, 9, 10, 16-18, is to nucleotide sequences, amino acid sequences and proteins with a zinc finger domain.

Invention 2, defined by claims 4, 5, 11, 12, 19-21, is to nucleotide sequences and amino acid sequences and proteins which are guanine exchange factors.

Invention 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins which are heat shock proteins or heat shock binding proteins.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☒ No protest accompanied the payment of additional search fees.

**INTERNATIONAL SEARCH REPORT**

International Application No.

**PCT/AU 98/00380**

<b>C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EMBL AC AF012106 DT 6 November 1997 Lloyd S E and Thakker R V DE Homo Sapiens DnaJ protein (HSPF <sub>2</sub> )mRNA, complete cds	1,6-8,13-15,22-24
P,X	EMBL AC AF 036875 DT 20 May 1998 Silins G, Grimmond S, Hayward N DE Mus musculus multiple endocrine neoplasia type I candidate protein number 18 mRNA, complete cds	1,6-8,13-15,22-24

**THIS PAGE BLANK (USP 10)**